GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 4488.31 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 901
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	901	100.0	220384	14	AC078926	AC078926 Homo sapi
	2	899.4	99.8	8368	8	AY438977	AY438977 Homo sapi
	3	899.4	99.8	13612	6	AX377239	AX377239 Sequence
c	4	899.4	99.8	122302	8	AC003982	AC003982 Homo sapi
	5	897.8	99.6	189729	14	AC073930	AC073930 Homo sapi
	6	337.2	37.4	173950	8	AL390961	AL390961 Human DNA
c	7	337.2	37.4	174937	14	AC022245	AC022245 Homo sapi
c	8	337.2	37.4	197245	14	AL591164	AL591164 Homo sapi
c	9	337.2	37.4	266738	14	AL138896	AL138896 Homo sapi
c	10	324.8	36.0	100506	14	AP002020	AP002020 Homo sapi
c	11	324.8	36.0	118642	8	AC126283	AC126283 Homo sapi
	12	324.2	36.0	161306	14	AC091056	AC091056 Homo sapi
	13	324.2	36.0	162100	8	AC018868	AC018868 Homo sapi
c	14	324.2	36.0	180230	14	AC090783	AC090783 Homo sapi
c	15	323.8	35.9	154592	8	AL591215	AL591215 Human DNA
c	16	323.8	35.9	179176	14	AC022857	AC022857 Homo sapi
	17	322.8	35.8	176438	14	AC126761	AC126761 Homo sapi
	18	322.8	35.8	181265	14	AC083994	AC083994 Homo sapi
	19	322.8	35.8	184777	8	AC026749	AC026749 Homo sapi
c	20	322	35.7	89673	8	AC004389	AC004389 Homo Sapi
	21	321.8	35.7	179993	14	AC015879	AC015879 Homo sapi
	22	321.2	35.6	158158	14	AC034304	AC034304 Homo sapi
	23	320	35.5	155405	8	AC009812	AC009812 Homo sapi
c	24	319.8	35.5	146181	8	AP001269	AP001269 Homo sapi
c	25	319.6	35.5	123288	8	AL354712	AL354712 Human DNA
	26	318.8	35.4	190626	14	AC062036	AC062036 Homo sapi
	27	318.6	35.4	112646	8	AC010240	AC010240 Homo sapi
c	28	318	35.3	166258	14	AL451066	AL451066 Homo sapi
	29	317.8	35.3	134341	8	AC099794	AC099794 Homo sapi
c	30	317.6	35.2	88546	8	AC008435	AC008435 Homo sapi
c	31	317.2	35.2	37490	8	AC026356	AC026356 Homo sapi
c	32	317.2	35.2	220480	14	AC023973	AC023973 Homo sapi
	33	317	35.2	167478	14	AC011148	AC011148 Homo sapi
c	34	317	35.2	176539	8	AC074213	AC074213 Homo sapi
	35	317	35.2	182103	14	AC016998	AC016998 Homo sapi
	36	316.8	35.2	167863	14	AL590651	AL590651 Homo sapi
c	37	316.6	35.1	171951	8	CNS05TDA	AL356020 Human chr
c	38	316.4	35.1	133512	8	AL590714	AL590714 Human DNA
c	39	315.6	35.0	157579	8	AC018639	AC018639 Human Chr
	40	315.6	35.0	159451	8	AC012618	AC012618 Homo sapi
c	41	315.6	35.0	199014	14	AC073849	AC073849 Homo sapi
	42	315.6	35.0	224573	8	AC008758	AC008758 Homo sapi
c	43	315.2	35.0	97916	8	HS408N23	Z98048 Human DNA s
c	44	315.2	35.0	290708	14	AL161641	AL161641 Homo sapi
c	45	314.8	34.9	161256	8	AC087855	AC087855 Homo sapi

ALIGNMENTS

RESULT 1
AC078926/c
LOCUS AC078926 220384 bp DNA linear HTG 26-MAR-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AC078926
VERSION AC078926.23 GI:19718703
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 220384)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,

```
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match           100.0%;  Score 901;  DB 14;  Length 220384;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-172;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69811 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 69752

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69751 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 69692

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69691 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 69632

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69631 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 69572

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69571 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 69512

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69511 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 69452

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69451 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 69392

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69391 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 69332

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      69331 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 69272

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69271 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 69212

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69211 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 69152

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69151 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 69092

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69091 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 69032

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69031 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 68972

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68971 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 68912

Qy        901 C 901
              |
Db      68911 C 68911
```

```
RESULT 2
AY438977
LOCUS       AY438977               8368 bp    DNA     linear   PRI 29-OCT-2003
DEFINITION  Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene,
            complete cds.
ACCESSION   AY438977
VERSION     AY438977.1  GI:37953284
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 8368)
  AUTHORS   Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W.,
            Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D.,
            Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2003) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use:  NIEHS-SNPs, Environmental Genome
            Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
            (URL: http://egp.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .8368
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   101. .257
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     variation       205
                     /frequency="0.01"
                     /replace="a"
     variation       306
                     /frequency="0.01"
                     /replace="a"
     repeat_region   326. .538
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     misc_feature    354. .496
                     /note="Region not scanned for variation"
     gene            875. .6548
                     /gene="PLA2G1B"
     mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                     /gene="PLA2G1B"
                     /product="phospholipase A2, group IB (pancreas)"
     CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                     /gene="PLA2G1B"
                     /codon_start=1
                     /product="phospholipase A2, group IB (pancreas)"
                     /protein_id="AAR05441.1"
                     /db_xref="GI:37953285"
                     /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                     NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                     AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
     repeat_region   1174. .1307
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       1286
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     repeat_region   1319. .1588
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       1343
                     /gene="PLA2G1B"
                     /frequency="0.03"
                     /replace="a"
     variation       1593
                     /gene="PLA2G1B"
                     /frequency="0.02"
                     /replace="t"
     repeat_region   1605. .1669
                     /rpt_family="L2"
                     /rpt_type=dispersed
     misc_feature    1638. .2188
                     /gene="PLA2G1B"
                     /note="Region not scanned for variation"
     repeat_region   1670. .1975
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   1976. .2038
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   2086. .2165
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   2194. .2319
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     misc_feature    2319. .2453
                     /gene="PLA2G1B"
                     /note="Region not scanned for variation"
     variation       3027
                     /gene="PLA2G1B"
                     /frequency="0.25"
                     /replace="t"
     repeat_region   3165. .3431
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       3262
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="t"
     variation       3321
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       3396
                     /gene="PLA2G1B"
                     /frequency="0.03"
                     /replace="g"
     variation       3630
```

```
                     /gene="PLA2G1B"
                     /frequency="0.10"
                     /replace="c"
     variation       3702
                     /gene="PLA2G1B"
                     /frequency="0.15"
                     /replace="a"
     variation       3768
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="t"
     variation       3818
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     repeat_region   3846. .3964
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   3977. .4280
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   4281. .4346
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   4347. .4634
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       4386
                     /gene="PLA2G1B"
                     /frequency="0.08"
                     /replace="t"
     repeat_region   4635. .4723
                     /rpt_family="L2"
                     /rpt_type=dispersed
     variation       5013
                     /gene="PLA2G1B"
                     /frequency="0.10"
                     /replace="a"
     repeat_region   5103. .5404
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       5133
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="c"
     variation       5543
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="c"
     variation       5555
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="t"
     variation       5556
                     /gene="PLA2G1B"
                     /frequency="0.15"
                     /replace="g"
     repeat_region   5577. .5629
                     /rpt_family="L1"
                     /rpt_type=dispersed
     variation       5580
                     /gene="PLA2G1B"
                     /frequency="0.09"
                     /replace="g"
     variation       5675
                     /gene="PLA2G1B"
                     /frequency="0.09"
                     /replace="g"
     repeat_region   5779. .6083
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       5929
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             99.8%;  Score 899.4;  DB 8;  Length 8368;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-171;
  Matches  900;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5074 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 5133

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5134 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 5193

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5194 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 5253

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5254 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 5313

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5314 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 5373

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5374 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 5433

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5434 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 5493

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5494 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 5553

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5554 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 5613

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5614 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 5673

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5674 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 5733

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5734 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 5793

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5794 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 5853

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5854 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 5913

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       5914 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 5973

Qy        901 C 901
              |
Db       5974 C 5974
```

RESULT 3
AX377239
LOCUS AX377239 13612 bp DNA linear PAT 18-MAR-2002
DEFINITION Sequence 1 from Patent WO0212562.
ACCESSION AX377239
VERSION AX377239.1 GI:19573528
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kazemi,A., Kliem,S.E. and Koshy,B.
TITLE Haplotypes of the pla2g1b gene
JOURNAL Patent: WO 0212562-A 1 14-FEB-2002;
Genaissance Pharmaceuticals, Inc. (US)
FEATURES Location/Qualifiers
source 1. .13612
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
variation 3845
/note="PS1: polymorphic base G or A"
variation 3968
/note="PS2: polymorphic base A or G"
variation 6060
/note="PS2: polymorphic base A or G"
variation 6844
/note="PS4: polymorphic base G or A"
variation 9531
/note="PS5: polymorphic base G or A"
ORIGIN

Query Match 99.8%; Score 899.4; DB 6; Length 13612;
Best Local Similarity 99.9%; Pred. No. 2e-171;
Matches 900; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy           1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8216 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8275

Qy          61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8276 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8335

Qy         121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8336 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8395

Qy         181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8396 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8455

Qy         241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8456 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8515

Qy         301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8516 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 8575

Qy         361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8576 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 8635

Qy         421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8636 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 8695

Qy         481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8696 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 8755

Qy         541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8756 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 8815

Qy         601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8816 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 8875

Qy         661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8876 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 8935

Qy         721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8936 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 8995

Qy         781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8996 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 9055

Qy         841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        9056 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 9115

Qy         901 C 901
               |
Db        9116 C 9116
```

RESULT 4
AC003982/c
LOCUS AC003982 122302 bp DNA linear PRI 18-MAR-1999
DEFINITION Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION AC003982
VERSION AC003982.1 GI:2769695
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 122302)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 122302)
AUTHORS Bradshaw,H., Wu,X. and Ozersky,P.
TITLE The sequence of Homo sapiens PAC clone 166H1
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 122302)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (13-JAN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 122302)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY: WUGSC

Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

```
FEATURES             Location/Qualifiers
     source          1. .122302
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q"
                     /clone="166H1"
                     /clone_lib="RPCI-1"
     repeat_region   1. .238
                     /rpt_family="Alu"
     repeat_region   239. .265
                     /rpt_family="AT_rich"
     repeat_region   268. .565
                     /rpt_family="Alu"
     repeat_region   1001. .1309
                     /rpt_family="Alu"
     repeat_region   1440. .1745
                     /rpt_family="Alu"
     repeat_region   1909. .1940
                     /rpt_family="MER1_type"
     repeat_region   1955. .1983
                     /rpt_family="AT_rich"
     repeat_region   2063. .2362
                     /rpt_family="Alu"
     repeat_region   2516. .2650
                     /rpt_family="Alu"
     repeat_region   2674. .2730
                     /rpt_family="L1"
     repeat_region   2735. .2788
                     /rpt_family="(GAAAA)n"
     repeat_region   2790. .3077
                     /rpt_family="Alu"
     repeat_region   3081. .3139
                     /rpt_family="L1"
     repeat_region   3716. .4020
                     /rpt_family="Alu"
     repeat_region   4196. .4492
                     /rpt_family="Alu"
     gene            4560. .14034
                     /gene="WUGSC:H_166H1.1"
     CDS             join(4560. .5056,13418. .13748,13882. .14034)
                     /gene="WUGSC:H_166H1.1"
                     /note="unknown function; 60% similar to Z50177
                     (PID:g927403) (PID:g927402); H_166H1.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB95634.1"
                     /db_xref="GI:2769696"
                     /translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
                     KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
                     QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
                     HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
                     LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
                     VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
     repeat_region   5366. .5655
                     /rpt_family="Alu"
     repeat_region   5672. .5798
                     /rpt_family="Alu"
     repeat_region   5802. .5834
                     /rpt_family="(TA)n"
     repeat_region   5834. .5857
                     /rpt_family="POLY_A"
     repeat_region   5870. .6184
                     /rpt_family="Alu"
     repeat_region   6224. .6268
                     /rpt_family="(CA)n"
     repeat_region   6269. .6558
                     /rpt_family="Alu"
     repeat_region   6729. .7029
                     /rpt_family="Alu"
     repeat_region   7051. .7172
                     /rpt_family="Alu"
     repeat_region   7192. .7500
                     /rpt_family="Alu"
     repeat_region   7518. .7820
                     /rpt_family="Alu"
     repeat_region   8207. .8500
                     /rpt_family="Alu"
     repeat_region   8576. .8878
                     /rpt_family="Alu"
     repeat_region   9023. .9154
                     /rpt_family="Alu"
     repeat_region   9165. .9470
                     /rpt_family="Alu"
     repeat_region   9887. .10188
                     /rpt_family="Alu"
     repeat_region   10221. .10342
                     /rpt_family="L1"
     repeat_region   10408. .10707
                     /rpt_family="Alu"
     repeat_region   10713. .11014
                     /rpt_family="Alu"
     repeat_region   11020. .11110
                     /rpt_family="Alu"
     repeat_region   11146. .11386
                     /rpt_family="Alu"
     repeat_region   11208. .11386
                     /rpt_family="Alu"
     repeat_region   11391. .11515
                     /rpt_family="Alu"
     repeat_region   11568. .11705
                     /rpt_family="Alu"
```

```
     repeat_region   11737. .12032
                     /rpt_family="Alu"
     repeat_region   12034. .12066
                     /rpt_family="7SLRNA"
     repeat_region   12075. .12372
                     /rpt_family="Alu"
     repeat_region   12448. .12753
                     /rpt_family="Alu"
     repeat_region   12763. .12874
                     /rpt_family="L2"
     repeat_region   12999. .13303
                     /rpt_family="Alu"
     misc_feature    complement(13674. .13752)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     misc_feature    complement(13840. .14195)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     repeat_region   14196. .14229
                     /rpt_family="AT_rich"
     repeat_region   14240. .14339
                     /rpt_family="U6"
     repeat_region   14349. .14646
                     /rpt_family="Alu"
     repeat_region   14652. .14940
                     /rpt_family="Alu"

  Query Match             99.8%;  Score 899.4;  DB 8;  Length 122302;
  Best Local Similarity   99.9%;  Pred. No. 1.4e-171;
  Matches  900;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24588 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 24529

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24528 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 24469

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24468 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 24409

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24408 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 24349

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24348 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 24289

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24288 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 24229

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24228 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 24169

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24168 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 24109

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24108 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 24049

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24048 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 23989

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23988 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 23929

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23928 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 23869

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23868 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 23809

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23808 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 23749

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      23748 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 23689

Qy        901 C 901
              |
Db      23688 C 23688


RESULT 5
AC073930
LOCUS       AC073930              189729 bp    DNA     linear   HTG 01-MAR-2002
DEFINITION  Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC073930
VERSION     AC073930.23  GI:18958593
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 189729)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
```

```
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14861669.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBWS
                Center clone name: RP11-144B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191640 bases at least Q40
                Consensus quality: 195135 bases at least Q30
                Consensus quality: 197786 bases at least Q20
                Estimated insert size: 193340; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64773: contig of 64773 bp in length
            *    64774    64873: gap of unknown length
            *    64874    99097: contig of 34224 bp in length
            *    99098    99197: gap of unknown length
            *    99198   135331: contig of 36134 bp in length
            *   135332   135431: gap of unknown length
            *   135432   161344: contig of 25913 bp in length
            *   161345   161444: gap of unknown length
            *   161445   171231: contig of 9787 bp in length
            *   171232   171331: gap of unknown length
            *   171332   180922: contig of 9591 bp in length
            *   180923   181022: gap of unknown length
            *   181023   187271: contig of 6249 bp in length
            *   187272   187371: gap of unknown length
            *   187372   189729: contig of 2358 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-144B2"
     gap             64774. .64873
                     /estimated_length=unknown
     gap             99098. .99197
                     /estimated_length=unknown
     gap             135332. .135431
                     /estimated_length=unknown
     gap             161345. .161444
                     /estimated_length=unknown
     gap             171232. .171331
                     /estimated_length=unknown
     gap             180923. .181022
                     /estimated_length=unknown
     gap             187272. .187371
                     /estimated_length=unknown
ORIGIN

  Query Match             99.6%;  Score 897.8;  DB 14;  Length 189729;
  Best Local Similarity   99.8%;  Pred. No. 2.8e-171;
  Matches  899;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172643 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 172702

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172703 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 172762

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172763 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 172822

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172823 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 172882

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172883 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 172942

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172943 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 173002

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173003 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 173062

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173063 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 173122

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173123 ACGCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 173182

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173183 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 173242

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173243 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 173302

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173303 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 173362

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173363 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 173422

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173423 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 173482

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
```

```
                    ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      173483 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 173542

Qy         901 C 901
                |
Db      173543 C 173543


RESULT 6
AL390961
LOCUS       AL390961               173950 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP13-16H11 on chromosome 10 Contains
            three novel genes, the gene for trans-prenyltransferase (TPT), the
            3' end of the SSH3BP gene for spectrin SH3 domain binding protein
            1, a heat shock protein pseudogene and a CpG island, complete
            sequence.
ACCESSION   AL390961
VERSION     AL390961.17  GI:19351892
KEYWORDS    HTG; CpG island; SSH3BP; TPT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 173950)
  AUTHORS   Lawlor,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 11, 2002 this sequence version replaced gi:18121480.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP13-16H11 is from the library RPCI-13.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .173950
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP13-16H11"
                     /clone_lib="RPCI-13.1"
     misc_feature    2000
                     /note="Clone_right_end: RP11-128B16"
     gene            join(32186. .32554,33872. .33974)
                     /locus_tag="RP13-16H11.7-001"
     mRNA            join(32186. .32554,33872. .33974)
                     /locus_tag="RP13-16H11.7-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: AA888857.1"
     gene            join(36944. .37240,37909. .38008,38349. .38551,39512. .39749,
                     42023. .42404)
                     /locus_tag="RP13-16H11.1-001"
     mRNA            join(36944. .37240,37909. .38008,38349. .38551,39512. .39749,
                     42023. .42404)
                     /locus_tag="RP13-16H11.1-001"
                     /product="novel transcript"
                     /note="match: ESTs: AA429709.1 AL042653.1 AL043087.2"
     gene            join(37058. .37240,38349. .38551,39512. .39615,42023. .42412)
                     /locus_tag="RP13-16H11.1-002"
     mRNA            join(37058. .37240,38349. .38551,39512. .39615,42023. .42412)
                     /locus_tag="RP13-16H11.1-002"
                     /product="novel transcript"
                     /note="match: ESTs: AI183567.1 AL040562.1"
     gene            join(38290. .38551,39512. .39615,42023. .42069)
                     /locus_tag="RP13-16H11.1-003"
     mRNA            join(38290. .38551,39512. .39615,42023. .42069)
                     /locus_tag="RP13-16H11.1-003"
                     /product="novel transcript"
                     /note="match: ESTs: AA412640.1"
     polyA_signal    42384. .42389
                     /locus_tag="RP13-16H11.1-002"
     polyA_site      42407
                     /locus_tag="RP13-16H11.1-002"
     gene            complement(join(53164. .53350,57365. .57452,65074. .65117))
                     /locus_tag="RP13-16H11.2-001"
     mRNA            complement(join(53164. .53350,57365. .57452,65074. .65117))
                     /locus_tag="RP13-16H11.2-001"
                     /product="novel transcript"
                     /note="match: ESTs: BG460048.1 BG460049.1"
     gene            join(86617. .86798,91120. .91152,93635. .93699,94244. .94352,
                     98596. .98726,109176. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,131455. .131535,
                     135291. .135756)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-001"
     mRNA            join(86617. .86798,91120. .91152,93635. .93699,94244. .94352,
                     98596. .98726,109176. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,131455. .131535,
                     135291. .135756)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-001"
                     /product="trans-prenyltransferase (TPT)"
                     /note="match: ESTs: AA889371.1 AL558274.1 BF982221.1
                     BI255737.1 BI761095.1 BM824056.1 BM994354.1 BQ018783.1
                     BQ212558.1
                     match: cDNAs: AF118395.1 AF118855.1 AK012307.1 AK024802.1
                     BC026820.1"
     CDS             join(86670. .86798,91120. .91152,93635. .93699,94244. .94352,
                     98596. .98726,109176. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,131455. .131535,
                     135291. .135431)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-001"
                     /standard_name="OTTHUMP00000019346"
                     /note="match: proteins: Q9CZQ1 Q9WU69 Q9Y2W5"
                     /codon_start=1
                     /product="trans-prenyltransferase (TPT)"
                     /protein_id="CAI17280.1"
                     /db_xref="GI:55959523"
                     /db_xref="GOA:Q5T2R2"
                     /db_xref="InterPro:IPR000092"
                     /db_xref="UniProt/TrEMBL:Q5T2R2"
                     /translation="MASRWWRWRRGCSWKPAARSPGPGSPGRAGPLGPSAAAEVRAQV
                     HRRKGLDLSQIPYINLVKHLTSACPNVCRISRFHHTTPDSKTHSGEKYTDPFKLGWRD
                     LKGLYEDIRKELLISTSELKEMSEYYFDGKGKAFRPIIVALMARACNIHHNNSRHVQA
                     SQRAIALIAEMIHTASLVHDDVIDDASSRRGKHTVNKIWGEKKAVLAGDLILSAASIA
                     LARIGNTTVISILTQVIEDLVRGEFLQLGSKENENERFAHYLEKTFKKTASLIANSCK
                     AVSVLGCPDPVVHEIAYQYGKNVGIAFQLIDDVLDFTSCSDQMGKPTSADLKLGLATG
                     PVLFACQQFPEMNAMIMRRFSLPGDVDRARQYVLQSDGVQQTTYLAQQYCHEAIREIS
                     KLRPSPERDALIQLSEIVLTRDK"
     gene            join(94266. .94352,98596. .98726,109093. .109317,
                     112764. .112875,112972. .112977)
```

```
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-004"
     mRNA            join(94266. .94352,98596. .98726,109093. .109317,
                     112764. .112875,112972. .112977)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-004"
                     /product="trans-prenyltransferase (TPT)"
                     /note="match: ESTs: AA362710.1 AI873444.1 BI062674.1"
     gene            join(103580. .103675,103811. .104363,104660. .104790)
                     /locus_tag="RP13-16H11.6-001"
                     /pseudo
     CDS             join(103580. .103675,103811. .104363,104660. .104790)
                     /locus_tag="RP13-16H11.6-001"
                     /note="match: proteins: AAC84170 AAH35665 O62564 O88686
                     O93240 P02827 P16627 P17066 P17879 P29843 P55063 P79984
                     Q05944 Q28222 Q8UWM6 Q8WQ17 Q91233 Q925V6 Q9NGK9 Q9QWJ5
                     Q9TUG3 Q9XZJ2"
                     /pseudo
                     /codon_start=1
                     /product="heat shock protein pseudogene"
     gene            join(106593. .107031,107691. .107736)
                     /locus_tag="RP13-16H11.5-001"
     mRNA            join(106593. .107031,107691. .107736)
                     /locus_tag="RP13-16H11.5-001"
                     /product="putative novel transcript"
                     /note="match: ESTs: BF387222.1"
     gene            join(<109301. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,129603. .129775,
                     131455. .131535,135291. .135499)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-002"
     mRNA            join(<109301. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,129603. .129775,
                     131455. .131535,135291. .135499)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-002"
                     /product="trans-prenyltransferase (TPT)"
                     /note="match: ESTs: BM806663.1"
     CDS             join(<109301. .109317,112764. .112875,112972. .113081,
                     124198. .124278,124424. .124537,129603. .129644)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-002"
                     /standard_name="OTTHUMP00000019347"
                     /codon_start=3
                     /product="trans-prenyltransferase (TPT)"
                     /protein_id="CAI17282.1"
                     /db_xref="GI:55959524"
                     /db_xref="GOA:Q5T2R1"
                     /db_xref="InterPro:IPR000092"
                     /db_xref="UniProt/TrEMBL:Q5T2R1"
                     /translation="WGEKKAVLAGDLILSAASIALARIGNTTVISILTQVIEDLVRGE
                     FLQLGSKENENERFAHYLEKTFKKTASLIANSCKAVSVLGCPDPVVHEIAYQYGKNVG
                     IAFQLIDDVLDFTSCSDQMGKPTSADLKLGLATGPVLFACQQVHCQGKTGSCCKH"
     gene            join(123609. .123681,124198. .124278,124424. .124537,
                     131455. .131535,135291. .135740)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-003"
     mRNA            join(123609. .123681,124198. .124278,124424. .124537,
                     131455. .131535,135291. .135740)
                     /gene="RP13-16H11.3"
                     /locus_tag="RP13-16H11.3-003"
                     /product="trans-prenyltransferase (TPT)"
                     /note="match: ESTs: BI756998.1"
     gene            join(complement(AL139404.9:77755. .77994),
```

```
  Query Match             37.4%;  Score 337.2;  DB 8;  Length 173950;
  Best Local Similarity   63.8%;  Pred. No. 2.8e-58;
  Matches  559;  Conservative    0;  Mismatches  313;  Indels    4;  Gaps    3;

Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||| |||||| |||| |||||| | ||| | ||||| || ||||||||||||||||||
Db     133844 GGGCGCAGTGGCTCACACCTGTAATCGCAGCATTTTGGGAGGCCAAGGTGGGTGGATCAC 133903

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
              ||||||| || |||| |||||||||||||| ||| ||||||||||||  ||||||||||
Db     133904 TTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAA 133963

Qy        150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 209
              ||||||| ||||||  || || ||  |||| |||| || ||| |||||||  ||||||||
Db     133964 ATACAAACATTAGCCAGGCGTGGTAGTGCACGCCTATAATCCTAGCTACTTAGGAGGCTG 134023

Qy        210 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCAC 269
              ||||| ||||||  ||||| |||||||||||||||||||||||||| ||||| | ||||
Db     134024 AGGCACGAGAATCCCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACGCCAC 134083

Qy        270 TGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 329
              |||| |||||||||||||||| ||| ||| || |||||||||||     ||| |||  |
Db     134084 TGCATTCCAGCCTGGGCGACAGAGCAAGATTCCATCTCAAAAAAAATTTAAAAAAAATAA 134143

Qy        330 AAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAG-ATACGTAAAT 388
              |  ||   |||| |  |          ||      ||   ||  | || ||| |
Db     134144 ATAAATGTCAGAGGGATCATGAGCCATGGAGAAGGAAATAGAATTATAGTATATGACTTG 134203

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
                | || ||||||  ||        |  |    |  |||  |   ||||| | || ||
Db     134204 ATTCTGCAATAAATATTTTTACATCATAGAAATGATACTTTATTAGTATTTGGCTTTTAG 134263

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
              ||  || |||       |  |        |  ||     |  ||||||||||  |  ||
Db     134264 AAAGAACCAATATTTTAAAACATGGTTATAGAATAGGATGTAAATATCTATAGTCTAACA 134323

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
                |  |  || |          |     |  || || |   |      ||| |
Db     134324 AACAGTATAGTTAAAGGAAAGGACAAGAGCTCATAGAAGATAAAATCTGAAGTTGATGCT 134383

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
                  | ||        |    || |   |  ||||  |      |
Db     134384 ACCTAACAGAAGTAATCAACAAAACAACTAAAGACAGCAGTGTGATGATGAGGCGGCACA 134443

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
              || |  ||          | |   | | | |       | | |  | |  || |
Db     134444 TTGTGGGTGGGCTAAACCCTCATCTTAACAAAAAGGTAACAGGTAATAACTAAAATTGGT 134503

Qy        689 --TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTG 745
                | ||   | | |||||||| |||||| |||||||||| ||||| ||||||||||||||
Db     134504 AATCAAGAAATAAAAACTAGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTG 134563

Qy        746 GGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATG 805
              ||||||| |||||     |||||||||||| | |||||| |||||||||||||| |||||||
Db     134564 GGAGGCCAAGGCAGGCAGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATG 134623

Qy        806 ATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGT 865
              |  |||| || |||||| |||||||||||||||||||| |||||||||||||| || |||||
Db     134624 ACAAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGCGCCTGT 134683

Qy        866 AGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              |||||||||||||  |||| ||||||||||  |||
Db     134684 AGTCCCAGCTACTTGGGAGGCTGAGGCAGGACAATC 134719
```

```
RESULT 7
AC022245/c
LOCUS       AC022245              174937 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-356G5 map 10, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC022245
VERSION     AC022245.3  GI:7230165
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 174937)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 10, clone RP11-356G5
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174937)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 174937)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6850466.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4212
                Center clone name: 356_G_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167019 bases at least Q40
                Consensus quality: 170472 bases at least Q30
                Consensus quality: 171999 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 173737; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1785: contig of 1785 bp in length
            *     1786     1885: gap of 100 bp
            *     1886     3111: contig of 1226 bp in length
            *     3112     3211: gap of 100 bp
            *     3212     5622: contig of 2411 bp in length
            *     5623     5722: gap of 100 bp
            *     5723    10591: contig of 4869 bp in length
            *    10592    10691: gap of 100 bp
            *    10692    15413: contig of 4722 bp in length
            *    15414    15513: gap of 100 bp
            *    15514    25047: contig of 9534 bp in length
            *    25048    25147: gap of 100 bp
            *    25148    36570: contig of 11423 bp in length
            *    36571    36670: gap of 100 bp
            *    36671    50263: contig of 13593 bp in length
            *    50264    50363: gap of 100 bp
            *    50364    64282: contig of 13919 bp in length
            *    64283    64382: gap of 100 bp
            *    64383    84772: contig of 20390 bp in length
            *    84773    84872: gap of 100 bp
            *    84873   105222: contig of 20350 bp in length
            *   105223   105322: gap of 100 bp
            *   105323   137522: contig of 32200 bp in length
            *   137523   137622: gap of 100 bp
            *   137623   174937: contig of 37315 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174937
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10"
                     /clone="RP11-356G5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1785
                     /note="assembly_fragment"
     gap             1786. .1885
                     /estimated_length=100
     misc_feature    1886. .3111
                     /note="assembly_fragment"
     gap             3112. .3211
                     /estimated_length=100
     misc_feature    3212. .5622
                     /note="assembly_fragment"
     gap             5623. .5722
                     /estimated_length=100
     misc_feature    5723. .10591
                     /note="assembly_fragment"
     gap             10592. .10691
                     /estimated_length=100
     misc_feature    10692. .15413
                     /note="assembly_fragment"
     gap             15414. .15513
                     /estimated_length=100
     misc_feature    15514. .25047
                     /note="assembly_fragment"
     gap             25048. .25147
                     /estimated_length=100
     misc_feature    25148. .36570
                     /note="assembly_fragment"
     gap             36571. .36670
                     /estimated_length=100
     misc_feature    36671. .50263
                     /note="assembly_fragment"
     gap             50264. .50363
                     /estimated_length=100
     misc_feature    50364. .64282
```

```
                    /note="assembly_fragment"
     gap            64283. .64382
                    /estimated_length=100
     misc_feature   64383. .84772
                    /note="assembly_fragment"
     gap            84773. .84872
                    /estimated_length=100
     misc_feature   84873. .105222
                    /note="assembly_fragment"
     gap            105223. .105322
                    /estimated_length=100
     misc_feature   105323. .137522
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
     gap            137523. .137622
                    /estimated_length=100
     misc_feature   137623. .174937
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
ORIGIN

  Query Match             37.4%;  Score 337.2;  DB 14;  Length 174937;
  Best Local Similarity   63.8%;  Pred. No. 2.8e-58;
  Matches  559;  Conservative    0;  Mismatches  313;  Indels    4;  Gaps    3;

Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||| ||||| |||| |||||||| | ||| | ||||| || ||||||||||||||||||
Db      50262 GGGCGCAGTGGCTCACACCTGTAATCGCAGCATTTTGGGAGGCCAAGGTGGGTGGATCAC 50203

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
              ||||||| || |||| |||||||||||||| ||| ||||||||||||  ||||||||||
Db      50202 TTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAA 50143

Qy        150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 209
              ||||||| ||||||   || || ||  |||| |||| || ||| |||||||  ||||||||
Db      50142 ATACAAACATTAGCCAGGCGTGGTAGTGCACGCCTATAATCCTAGCTACTTAGGAGGCTG 50083

Qy        210 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCAC 269
              ||||| ||||||  ||||| ||||||||||||||||||||||||||| ||||| | ||||
Db      50082 AGGCACGAGAATCCCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACGCCAC 50023

Qy        270 TGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 329
              |||| ||||||||||||||| ||| ||| || ||||||||||||     ||| |||  |
Db      50022 TGCATTCCAGCCTGGGCGACAGAGCAAGATTCCATCTCAAAAAAAATTTAAAAAAAATAA 49963

Qy        330 AAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAG-ATACGTAAAT 388
              |   ||   ||||  |  |          ||       ||   ||  | || ||| |
Db      49962 ATAAATGTCAGAGGGATCATGAGCCATGGAGAAGGAAATAGAATTATAGTATATGACTTG 49903

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
              | || ||||||  ||         |  |    |  |||  |   |||||  | || ||
Db      49902 ATTCTGCAATAAATATTTTTACATCATAGAAATGATACTTTATTAGTATTTGGCTTTTAG 49843

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
              ||  || |||      |  |         |  ||     |  ||||||||||  |  ||
Db      49842 AAAGAACCAATATTTTAAAACATGGTTATAGAATAGGATGTAAATATCTATAGTCTAACA 49783

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
                |    |  || |          |      |  || || |   |    |||  |
Db      49782 AACAGTATAGTTAAAGGAAAGGACAAGAGCTCATAGAAGATAAAATCTGAAGTTGATGCT 49723

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
                  | ||          |      || |   |  ||||  |      |
Db      49722 ACCTAACAGAAGTAATCAACAAAACAACTAAAGACAGCAGTGTGATGATGAGGCGGCACA 49663

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
              || |   ||          | |   | | |       |  | |  | |  || |
Db      49662 TTGTGGGTGGGCTAAACCCTCATCTTAACAAAAAGGTAACAGGTAATAACTAAAATTGGT 49603

Qy        689 --TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTG 745
                | ||   | | |||||||| |||||| |||||||||| ||||| ||||||||||||||
```

```
Db      49602 AATCAAGAAATAAAAACTAGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTG 49543

Qy        746 GGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATG 805
              ||||||| |||||    |||||||||||| | |||||| |||||||||||||| |||||||
Db      49542 GGAGGCCAAGGCAGGCAGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATG 49483

Qy        806 ATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGT 865
              |  |||| || |||||| |||||||||||||||||||| |||||||||||||| || |||||
Db      49482 ACAAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGCGCCTGT 49423

Qy        866 AGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              |||||||||||||  |||| ||||||||||  |||
Db      49422 AGTCCCAGCTACTTGGGAGGCTGAGGCAGGACAATC 49387
```

```
RESULT 8
AL591164/c
LOCUS       AL591164              197245 bp    DNA     linear   HTG 19-DEC-2001
DEFINITION  Homo sapiens chromosome 10 clone RP13-31H8, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AL591164
VERSION     AL591164.3  GI:17973970
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 20, 2001 this sequence version replaced gi:14575431.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bB31H8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 190732 bases at least Q40
            Consensus quality: 193432 bases at least Q30
            Consensus quality: 194855 bases at least Q20
            Insert size: 195945; sum-of-contigs
            Insert size: 204159; 4.1% error; agarose-fp
            Quality coverage: 4.66x in Q20 bases; sum-of-contigs Quality
            coverage: 4.47x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6607: contig of 6607 bp in length
            *     6608     6707: gap of 100 bp
            *     6708    13612: contig of 6905 bp in length
            *    13613    13712: gap of 100 bp
            *    13713    39295: contig of 25583 bp in length
            *    39296    39395: gap of 100 bp
            *    39396    52647: contig of 13252 bp in length
            *    52648    52747: gap of 100 bp
            *    52748    55062: contig of 2315 bp in length
            *    55063    55162: gap of 100 bp
            *    55163    73956: contig of 18794 bp in length
```

```
            *    73957    74056: gap of 100 bp
            *    74057    94138: contig of 20082 bp in length
            *    94139    94238: gap of 100 bp
            *    94239   118037: contig of 23799 bp in length
            *   118038   118137: gap of 100 bp
            *   118138   120149: contig of 2012 bp in length
            *   120150   120249: gap of 100 bp
            *   120250   174347: contig of 54098 bp in length
            *   174348   174447: gap of 100 bp
            *   174448   177356: contig of 2909 bp in length
            *   177357   177456: gap of 100 bp
            *   177457   180727: contig of 3271 bp in length
            *   180728   180827: gap of 100 bp
            *   180828   194321: contig of 13494 bp in length
            *   194322   194421: gap of 100 bp
            *   194422   197245: contig of 2824 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197245
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP13-31H8"
                     /clone_lib="RPCI-13.1"
     misc_feature    1. .6607
                     /note="assembly_fragment:00616
                     fragment_chain:1"
     misc_feature    6708. .13612
                     /note="assembly_fragment:01749
                     fragment_chain:1"
     misc_feature    13713. .39295
                     /note="assembly_fragment:00559
                     fragment_chain:1"
     misc_feature    39396. .52647
                     /note="assembly_fragment:01954
                     fragment_chain:1"
     misc_feature    52748. .55062
                     /note="assembly_fragment:00859
                     fragment_chain:1"
     misc_feature    55163. .73956
                     /note="assembly_fragment:01247
                     fragment_chain:1"
     misc_feature    74057. .94138
                     /note="assembly_fragment:01887
                     fragment_chain:1"
     misc_feature    94239. .118037
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    118138. .120149
                     /note="assembly_fragment:00714
                     fragment_chain:2"
     misc_feature    120250. .174347
                     /note="assembly_fragment:00818
                     fragment_chain:2"
     misc_feature    174448. .177356
                     /note="assembly_fragment:00335
                     fragment_chain:2"
     misc_feature    177457. .180727
                     /note="assembly_fragment:01862"
     misc_feature    180828. .194321
                     /note="assembly_fragment:01417
                     fragment_chain:3"
     misc_feature    194422. .197245
                     /note="assembly_fragment:01084
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             37.4%;  Score 337.2;  DB 14;  Length 197245;
  Best Local Similarity   63.8%;  Pred. No. 2.7e-58;
  Matches  559;  Conservative    0;  Mismatches  313;  Indels    4;  Gaps    3;

Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||| ||||| ||||  ||||||| | ||| | ||||| || ||||||||||||||||||
Db     115022 GGGCGCAGTGGCTCACACCTGTAATCGCAGCATTTTGGGAGGCCAAGGTGGGTGGATCAC 114963

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
              ||||||| || |||| |||||||||||| ||| ||||||||||||  ||||||||||||
Db     114962 TTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAA 114903

Qy        150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 209
              ||||||| ||||||  || || ||  |||| |||| || ||| |||||||  ||||||||
Db     114902 ATACAAACATTAGCCAGGCGTGGTAGTGCACGCCTATAATCCTAGCTACTTAGGAGGCTG 114843

Qy        210 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCAC 269
              |||||  ||||||  ||||| ||||||||||||||||||||||||||||| ||||| | ||||
Db     114842 AGGCACGAGAATCCCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACGCCAC 114783

Qy        270 TGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 329
              |||| |||||||||||||||| ||| ||| || |||||||||||      ||| ||| |
Db     114782 TGCATTCCAGCCTGGGCGACAGAGCAAGATTCCATCTCAAAAAAAATTTAAAAAAAATAA 114723

Qy        330 AAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAG-ATACGTAAAT 388
              |  ||   |||| |  |           ||      ||  ||  | || ||| |
Db     114722 ATAAATGTCAGAGGGATCATGAGCCATGGAGAAGGAAATAGAATTATAGTATATGACTTG 114663

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
                | || ||||||  ||          |  |    |  |||   |  ||||| | || ||
Db     114662 ATTCTGCAATAAATATTTTTACATCATAGAAATGATACTTTATTAGTATTTGGCTTTTAG 114603

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
              ||  || |||       |  |          |  ||    |  |||||||||||  |  ||
Db     114602 AAAGAACCAATATTTTAAAACATGGTTATAGAATAGGATGTAAATATCTATAGTCTAACA 114543

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
                |  |  || |          |    |  || || |   |     |||  |
Db     114542 AACAGTATAGTTAAAGGAAAGGACAAGAGCTCATAGAAGATAAAATCTGAAGTTGATGCT 114483

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
                  | ||          |      || |   |  ||||  |      |
Db     114482 ACCTAACAGAAGTAATCAACAAAACAACTAAAGACAGCAGTGTGATGATGAGGCGGCACA 114423

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
              || |   ||           | |   | | | |     |  | |  | |  || |
Db     114422 TTGTGGGTGGGCTAAACCCTCATCTTAACAAAAAGGTAACAGGTAATAACTAAAATTGGT 114363

Qy        689 --TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTG 745
                | ||   | | |||||||| |||||| |||||||||| ||||| ||||||||||||||
Db     114362 AATCAAGAAATAAAAACTAGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTG 114303

Qy        746 GGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATG 805
              |||||||  |||||    |||||||||||| | |||||| |||||||||||| |||||||
Db     114302 GGAGGCCAAGGCAGGCAGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATG 114243

Qy        806 ATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGT 865
              |  |||| || |||||| |||||||||||||||||| |||||||||||| || |||||
Db     114242 ACAAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGCGCCTGT 114183

Qy        866 AGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              |||||||||||||  ||||  ||||||||||  |||
Db     114182 AGTCCCAGCTACTTGGGAGGCTGAGGCAGGACAATC 114147


RESULT 9
AL138896/c
LOCUS       AL138896              266738 bp    DNA     linear   HTG 30-JAN-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-133M13, 16 unordered pieces.
ACCESSION   AL138896
VERSION     AL138896.11  GI:14625510
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 6, 2001 this sequence version replaced gi:14586010.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA133M13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 262635 bases at least Q40
            Consensus quality: 263809 bases at least Q30
            Consensus quality: 264520 bases at least Q20
            Insert size: 265238; sum-of-contigs
            Insert size: 141679; 7.5% error; agarose-fp
            Quality coverage: 10.63x in Q20 bases; sum-of-contigs Quality
            coverage: 20.65x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    32499: contig of 32499 bp in length
            *    32500    32599: gap of 100 bp
            *    32600    34912: contig of 2313 bp in length
            *    34913    35012: gap of 100 bp
            *    35013   106818: contig of 71806 bp in length
            *   106819   106918: gap of 100 bp
            *   106919   125886: contig of 18968 bp in length
            *   125887   125986: gap of 100 bp
            *   125987   145316: contig of 19330 bp in length
            *   145317   145416: gap of 100 bp
            *   145417   153004: contig of 7588 bp in length
            *   153005   153104: gap of 100 bp
            *   153105   155300: contig of 2196 bp in length
            *   155301   155400: gap of 100 bp
            *   155401   185898: contig of 30498 bp in length
            *   185899   185998: gap of 100 bp
            *   185999   188966: contig of 2968 bp in length
            *   188967   189066: gap of 100 bp
            *   189067   192900: contig of 3834 bp in length
            *   192901   193000: gap of 100 bp
            *   193001   197749: contig of 4749 bp in length
            *   197750   197849: gap of 100 bp
            *   197850   214650: contig of 16801 bp in length
            *   214651   214750: gap of 100 bp
            *   214751   218266: contig of 3516 bp in length
            *   218267   218366: gap of 100 bp
            *   218367   255095: contig of 36729 bp in length
            *   255096   255195: gap of 100 bp
            *   255196   264279: contig of 9084 bp in length
            *   264280   264379: gap of 100 bp
            *   264380   266738: contig of 2359 bp in length.
FEATURES             Location/Qualifiers
     source          1. .266738
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-133M13"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .32499
                     /note="assembly_fragment:02113
                     fragment_chain:1"
     misc_feature    32600. .34912
                     /note="assembly_fragment:02238
                     fragment_chain:1"
     misc_feature    35013. .106818
                     /note="assembly_fragment:07688
                     fragment_chain:2"
     misc_feature    106919. .125886
                     /note="assembly_fragment:03826
                     fragment_chain:2"
     misc_feature    125987. .145316
                     /note="assembly_fragment:07862
                     fragment_chain:3"
     misc_feature    145417. .153004
                     /note="assembly_fragment:01625
                     fragment_chain:3"
     misc_feature    153105. .155300
                     /note="assembly_fragment:01523"
     misc_feature    155401. .185898
                     /note="assembly_fragment:01783"
     misc_feature    185999. .188966
                     /note="assembly_fragment:03724"
     misc_feature    189067. .192900
                     /note="assembly_fragment:05406"
     misc_feature    193001. .197749
                     /note="assembly_fragment:05944"
     misc_feature    197850. .214650
                     /note="assembly_fragment:06359"
     misc_feature    214751. .218266
                     /note="assembly_fragment:06448"
     misc_feature    218367. .255095
                     /note="assembly_fragment:07266"
     misc_feature    255196. .264279
                     /note="assembly_fragment:07507"
     misc_feature    264380. .266738
                     /note="assembly_fragment:07874"
ORIGIN

  Query Match             37.4%;  Score 337.2;  DB 14;  Length 266738;
  Best Local Similarity   63.8%;  Pred. No. 2.6e-58;
  Matches  559;  Conservative    0;  Mismatches  313;  Indels    4;  Gaps    3;

Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||| ||||| |||| |||||||| | ||| | ||||| || ||||||||||||||||||
Db      86293 GGGCGCAGTGGCTCACACCTGTAATCGCAGCATTTTGGGAGGCCAAGGTGGGTGGATCAC 86234

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
              ||||||| || |||| |||||||||||||| ||| ||||||||||||  ||||||||||||
Db      86233 TTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAA 86174

Qy        150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 209
              ||||||| ||||||  || || ||  |||| |||| || ||| |||||||  ||||||||
Db      86173 ATACAAACATTAGCCAGGCGTGGTAGTGCACGCCTATAATCCTAGCTACTTAGGAGGCTG 86114

Qy        210 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCAC 269
              ||||| ||||||  ||||| |||||||||||||||||||||||||||| ||||| | ||||
Db      86113 AGGCACGAGAATCCCTTGAGCCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACGCCAC 86054

Qy        270 TGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 329
              |||| |||||||||||||||| ||| ||| || ||||||||||||     ||| |||  |
Db      86053 TGCATTCCAGCCTGGGCGACAGAGCAAGATTCCATCTCAAAAAAAATTTAAAAAAAATAA 85994

Qy        330 AAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAG-ATACGTAAAT 388
              |   ||  |||| | |         ||    ||   ||  | || ||| |
Db      85993 ATAAATGTCAGAGGGATCATGAGCCATGGAGAAGGAAATAGAATTATAGTATATGACTTG 85934

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
              | || ||||||  ||         |  |    | |||  |   ||||| | || ||
Db      85933 ATTCTGCAATAAATATTTTTACATCATAGAAATGATACTTTATTAGTATTTGGCTTTTAG 85874

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
```

```
Db      85873 AAAGAACCAATATTTTAAAACATGGTTATAGAATAGGATGTAAATATCTATAGTCTAACA 85814

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568

Db      85813 AACAGTATAGTTAAAGGAAAGGACAAGAGCTCATAGAAGATAAAATCTGAAGTTGATGCT 85754

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628

Db      85753 ACCTAACAGAAGTAATCAACAAAACAACTAAAGACAGCAGTGTGATGATGAGGCGGCACA 85694

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688

Db      85693 TTGTGGGTGGGCTAAACCCTCATCTTAACAAAAAGGTAACAGGTAATAACTAAAATTGGT 85634

Qy        689 --TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTG 745

Db      85633 AATCAAGAAATAAAAACTAGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTG 85574

Qy        746 GGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATG 805

Db      85573 GGAGGCCAAGGCAGGCAGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGCCAACATG 85514

Qy        806 ATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGT 865

Db      85513 ACAAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGCGCCTGT 85454

Qy        866 AGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901

Db      85453 AGTCCCAGCTACTTGGGAGGCTGAGGCAGGACAATC 85418


RESULT 10
AP002020/c
LOCUS       AP002020              100506 bp    DNA     linear   HTG 10-AUG-2000
DEFINITION  Homo sapiens chromosome 4 clone 2087A2 map 4q22-q24, *** SEQUENCING
            IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AP002020
VERSION     AP002020.2  GI:9795241
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 100506)
  AUTHORS   Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
            University, Institute of Genetics; 155 Li-Rong St. Section 2,
            Peitou, Taipei, Taiwan 11221, Republic of China
            (E-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
            Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT     On Aug 11, 2000 this sequence version replaced gi:7798580.
            These sequences are draft human sequences, not finished sequences.
            These sequences are unordered pieces. And gaps between the contigs
            of the same clone are represented as 100 N.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6503: contig of 6503 bp in length
            *     6504     6603: gap of 100 bp
            *     6604    15759: contig of 9156 bp in length
            *    15760    15859: gap of 100 bp
            *    15860    25981: contig of 10122 bp in length
            *    25982    26081: gap of 100 bp
            *    26082    38177: contig of 12096 bp in length
            *    38178    38277: gap of 100 bp
            *    38278    51292: contig of 13015 bp in length
            *    51293    51392: gap of 100 bp
            *    51393    62369: contig of 10977 bp in length
            *    62370    62469: gap of 100 bp
            *    62470   100506: contig of 38037 bp in length.
FEATURES             Location/Qualifiers
     source          1. .100506
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q22-q24"
                     /clone="2087A2"
ORIGIN

  Query Match           36.0%;  Score 324.8;  DB 14;  Length 100506;
  Best Local Similarity   63.7%;  Pred. No. 9.5e-56;
  Matches  575;  Conservative    0;  Mismatches  317;  Indels   10;  Gaps    5;

Qy          9 TTAATTAAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGA 68

Db      87376 TTTAAAATACCTTTGTAGGCGGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 87317

Qy         69 AGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGT 128

Db      87316 AGGCCAAGGCAGATGGATCACTTGAGATCAGGAGTTCGAGACCAGCCTGGCCAATACGGT 87257

Qy        129 GAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAG 188

Db      87256 GAAACCCCTTCTCTACTAAAAATACAAAAATCAGCCGGGCATGGTGGCACACGCCTGTAA 87197

Qy        189 TCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 248

Db      87196 TCGCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGC 87137

Qy        249 AGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCA 308

Db      87136 CGTGAGCCGAGATCACGCCACTGCACTCCAGCCTGGACGACAGAGCAAGGCTTGGTCTCA 87077

Qy        309 AAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAG 368

Db      87076 AAACAAACAAAAAACAAAATAAGTGA-GAGAGGAAACTTTACTTGTTCTTTTTCTTTTAG 87018

Qy        369 GTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTT 428

Db      87017 TTTAGTAAATGCTGCAATTTGATTTTAACGAATGTTGGTCTTTTTATGGGCCTATATAAG 86958

Qy        429 ATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTA--ATGCAACTTCAAACTCTA 486

Db      86957 TGTATTATTTTTGAAACAAAAAATGAAAATTACCTATTTAGTATGAAACCTTTGCTTATC 86898

Qy        487 CATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCA 546

Db      86897 TTTTGCTGTGTATTTGCTCAATTCTGAATTGAACTTGAGATTGTCACACACAGATTGCAT 86838

Qy        547 GTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTC 606

Db      86837 TTTCAACTGAAATGAGATGATTTTTGTCTTTACTCTTAAAACTTAGTGAGAAAAAGCTTG 86778

Qy        607 CATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTC-----CATTTAAGGA 661

Db      86777 CTTGAACAAGAAGTTTTAAAGTATGTCAAAATGTTTATTACTTTCTTAAAAATGTCAAGC 86718

Qy        662 TGTTCCATAATTTTTAAAATACATGCTTTAAAGTAG-AGAAACTAGGTTGGGCATGGTGG 720

Db      86717 TGCTCTTCTTTCTCAGAAATCCAAAACTTGTAGAAGCACAGATCAGGCCAGGTGCTGTGG 86658

Qy        721 CTCACGCCTGT-ATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGG 779

Db      86657 CTGACGCCTGTAATCCCAGCATTTTGGGAGGCTGAGGCAGGCAGATCACCTGTGGTCAGG 86598

Qy        780 AGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATT 839
```

```
Db      86597 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAACTT 86538

Qy        840 AGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTA 899
              ||| |||| |||||||  || |||||| ||| ||||| |  |||| |||||||||||| |
Db      86537 AGCCGGGCGTGGTGGCGGGCGCCTGTAATCCTAGCTATTTGGGAGGCTGAGGCAGGAGAA 86478

Qy        900 TC 901
              ||
Db      86477 TC 86476
```

RESULT 11
AC126283/c
LOCUS AC126283 118642 bp DNA linear PRI 10-AUG-2002
DEFINITION Homo sapiens BAC clone RP11-602N24 from 4, complete sequence.
ACCESSION AC126283
VERSION AC126283.3 GI:22091403
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118642)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 118642)
AUTHORS Yoakum,M., Kozlowicz,A. and Creason,K.
TITLE The sequence of Homo sapiens BAC clone RP11-602N24
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 118642)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 118642)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 118642)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 118642)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 2, 2002 this sequence version replaced gi:22002234.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0602N24
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC004067, 2000 bp overlap; the clone sequenced to the right is AC005509, 2000 bp overlap.

Polymorphisms exist between AC004067 and AC126283.

FEATURES	Location/Qualifiers
source	1. .118642
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-602N24"
	/clone_lib="RPCI-11"
repeat_region	56. .78
	/rpt_family="AT_rich"
repeat_region	109. .196
	/rpt_family="MIR"
repeat_region	310. .412
	/rpt_family="L2"
repeat_region	694. .775
	/rpt_family="Tip100"
repeat_region	776. .1073
	/rpt_family="Alu"
repeat_region	1074. .1160
	/rpt_family="Tip100"
repeat_region	1182. .1469
	/rpt_family="Alu"
repeat_region	1679. .2085
	/rpt_family="L2"
repeat_region	2086. .2390
	/rpt_family="Alu"
repeat_region	2391. .2719
	/rpt_family="L2"
repeat_region	2720. .3023
	/rpt_family="Alu"
repeat_region	3043. .3078
	/rpt_family="AT_rich"
repeat_region	3151. .3452
	/rpt_family="Alu"
repeat_region	3674. .3708
	/rpt_family="AT_rich"
repeat_region	3901. .4585
	/rpt_family="ERV1"
repeat_region	4975. .5569
	/rpt_family="L2"
repeat_region	5804. .6054
	/rpt_family="ERVL"
repeat_region	6306. .6609
	/rpt_family="Alu"
repeat_region	6685. .7017
	/rpt_family="MaLR"

```
     repeat_region   7123. .7308
                     /rpt_family="L1"
     repeat_region   7801. .8016
                     /rpt_family="L2"
     repeat_region   8210. .8595
                     /rpt_family="Alu"
     repeat_region   8635. .8790
                     /rpt_family="MIR"
     repeat_region   9622. .9842
                     /rpt_family="MIR"
     repeat_region   10227. .10353
                     /rpt_family="MIR"
     repeat_region   10372. .10473
                     /rpt_family="MIR"
     repeat_region   11326. .11395
                     /rpt_family="Alu"
     repeat_region   11727. .11815
                     /rpt_family="L2"
     repeat_region   11863. .11922
                     /rpt_family="L1"
     repeat_region   11923. .12193
                     /rpt_family="Alu"
     repeat_region   12194. .12393
                     /rpt_family="L1"
     repeat_region   12459. .12542
                     /rpt_family="L1"
     repeat_region   12979. .13072
                     /rpt_family="L1"
     repeat_region   13395. .13664
                     /rpt_family="Alu"
     repeat_region   14111. .14289
                     /rpt_family="AcHobo"
     repeat_region   14828. .15110
                     /rpt_family="CR1"
     repeat_region   15160. .15324
                     /rpt_family="L2"
     repeat_region   15371. .15654
                     /rpt_family="Alu"
     repeat_region   15914. .16246
                     /rpt_family="MER1_type"
     repeat_region   16461. .17021
                     /rpt_family="L2"
     repeat_region   18185. .18493
                     /rpt_family="Alu"
     repeat_region   18566. .18653
                     /rpt_family="MIR"
     repeat_region   19003. .19405
                     /rpt_family="ERVL"
     repeat_region   19637. .19938
                     /rpt_family="Alu"
     repeat_region   20076. .20150
                     /rpt_family="MIR"
     repeat_region   20528. .20558
                     /rpt_family="AT_rich"
     repeat_region   21061. .21745
                     /rpt_family="(TA)n"
     repeat_region   21748. .21838
                     /rpt_family="L1"
     repeat_region   21841. .22083
                     /rpt_family="(TTATA)n"
     repeat_region   22086. .22195
                     /rpt_family="(TA)n"
     repeat_region   22196. .22479
                     /rpt_family="Alu"
     repeat_region   23128. .23524
                     /rpt_family="L1"
     repeat_region   23599. .23977

  Query Match           36.0%;  Score 324.8;  DB 8;  Length 118642;
  Best Local Similarity  63.7%;  Pred. No. 9.2e-56;
  Matches  575;  Conservative    0;  Mismatches  317;  Indels   10;  Gaps    5;

Qy          9 TTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGA 68

Db      82886 TTTAAAATACCTTTGTAGGCGGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 82827

Qy         69 AGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGT 128

Db      82826 AGGCCAAGGCAGATGGATCACTTGAGATCAGGAGTTCGAGACCAGCCTGGCCAATACGGT 82767

Qy        129 GAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAG 188

Db      82766 GAAACCCCTTCTCTACTAAAAATACAAAAATCAGCCGGGCATGGTGGCACACGCCTGTAA 82707

Qy        189 TCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 248

Db      82706 TCGCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGC 82647

Qy        249 AGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCA 308

Db      82646 CGTGAGCCGAGATCACGCCACTGCACTCCAGCCTGGACGACAGAGCAAGGCTTGGTCTCA 82587

Qy        309 AAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAG 368

Db      82586 AAACAAACAAAAAACAAAATAAGTGA-GAGAGGAAACTTTACTTGTTCTTTTTCTTTTAG 82528

Qy        369 GTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTT 428

Db      82527 TTTAGTAAATGCTGCAATTTGATTTTAACGAATGTTGGTCTTTTTATGGGCCTATATAAG 82468

Qy        429 ATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTA--ATGCAACTTCAAACTCTA 486

Db      82467 TGTATTATTTTTGAAACAAAAAATGAAAATTACCTATTTAGTATGAAACCTTTGCTTATC 82408

Qy        487 CATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCA 546

Db      82407 TTTTGCTGTGTATTTGCTCAATTCTGAATTGAACTTGAGATTGTCACACACAGATTGCAT 82348

Qy        547 GTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTC 606

Db      82347 TTTCAACTGAAATGAGATGATTTTTGTCTTTACTCTTAAAACTTAGTGAGAAAAAGCTTG 82288

Qy        607 CATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTC-----CATTTAAGGA 661

Db      82287 CTTGAACAAGAAGTTTTAAAGTATGTCAAAATGTTTATTACTTTCTTAAAAATGTCAAGC 82228

Qy        662 TGTTCCATAATTTTTAAAATACATGCTTTAAAGTAG-AGAAACTAGGTTGGGCATGGTGG 720

Db      82227 TGCTCTTCTTTCTCAGAAATCCAAAACTTGTAGAAGCACAGATCAGGCCAGGTGCTGTGG 82168

Qy        721 CTCACGCCTGT-ATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGG 779

Db      82167 CTGACGCCTGTAATCCCAGCATTTTGGGAGGCTGAGGCAGGCAGATCACCTGTGGTCAGG 82108

Qy        780 AGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATT 839

Db      82107 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAACTT 82048

Qy        840 AGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTA 899

Db      82047 AGCCGGGCGTGGTGGCGGGCGCCTGTAATCCTAGCTATTTGGGAGGCTGAGGCAGGAGAA 81988

Qy        900 TC 901

Db      81987 TC 81986


RESULT 12
AC091056
LOCUS       AC091056              161306 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens chromosome 15 clone RP11-83J16 map 15, 4 unordered
            pieces.
ACCESSION   AC091056
VERSION     AC091056.4  GI:17136126
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 161306)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-83J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161306)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161306)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 28, 2001 this sequence version replaced gi:15706088.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11893
                Center clone name: 83_J_16
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    24848: contig of 24848 bp in length
            *   24849    24948: gap of 100 bp
            *   24949   132812: contig of 107864 bp in length
            *  132813   132912: gap of 100 bp
            *  132913   135523: contig of 2611 bp in length
            *  135524   135623: gap of 100 bp
            *  135624   161306: contig of 25683 bp in length.
FEATURES            Location/Qualifiers
     source         1. .161306
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="15"
                    /map="15"
                    /clone="RP11-83J16"
                    /clone_lib="RPCI-11 Human Male BAC"
     gap            24849. .24948
                    /estimated_length=100
     gap            132813. .132912
                    /estimated_length=100
     gap            135524. .135623
                    /estimated_length=100
ORIGIN

  Query Match           36.0%;  Score 324.2;  DB 14;  Length 161306;
  Best Local Similarity  62.7%;  Pred. No. 1.2e-55;
  Matches  564;  Conservative    0;  Mismatches  313;  Indels   22;  Gaps    3;

Qy          4 TCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 63
              ||| ||      || ||    |     |||  ||||| |||| |||| ||| | ||| |||
Db     105318 TCAAATATGAAAATAAGTGTTACACCGGGTGCAGTGGCTCATGCCTTTAATCCCAGCACT 105377

Qy         64 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 123
              |||| || |  | |  |||||||||| |||||||||| ||||| |||| ||||||  |||
Db     105378 TTGGGAGGCTGAAGCAGGTGGATCACCTGAGGTGAGGAGTTCAAGACTAGCCTGACCAAC 105437

Qy        124 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 183
              |||||||||||| ||||||||||||||||||||||||||||||| || ||| ||||  ||
Db     105438 ATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGCACACC 105497

Qy        184 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 243
              |||| |||||||||||| |||||||||||||||||||||||||||| ||||||| ||
Db     105498 TGTAATCCCAGCTACTCAGGAGGCTGAGGCAAGAGAATTGCTTGAACCCGGGAGGCGGAA 105557

Qy        244 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 303
              |||||||||| |||||||  ||||| |||||||||||||||| |||| || |||||||  |
Db     105558 GTTGCAGTGAACCGAGATTGCACCATTGCACTCCAGCCTGGGGGACAGAGTGAGACTCCA 105617

Qy        304 TCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCC 363
              |||||||||| ||| |||  ||| |||        | |        | |||||||
Db     105618 TCTCAAAAAAAAAAAAAAAAGAAAGAAAAAGAAAAATAAGTGTTTATCTAATAA------ 105671

Qy        364 TGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTA 423
                      | ||      |     | | | || || | |     | || | || |
Db     105672 ------------TGCTAAAATGACATATTCTTAAAATGTATCACATTCTCTTCAGTGAA 105718

Qy        424 ATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACT 483
                    |  ||||        || |    ||  | | | |||
Db     105719 GATAGTATAATTATAGAAAAAATCATGATATCTAAACAATATGTATGTTTTAACTTTTTG 105778

Qy        484 CTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAA 543
                |  || |   || ||   ||     ||| |      ||||||       |  ||  |
Db     105779 ATCATAAATTGAATAGTAGAAGAATCCAAAGCCTGGGGAAATCCTTTAGGGAAGGGAGGA 105838

Qy        544 CCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAAT 603
                  ||||| |  |  ||   | |      || |     |||| || | | ||||
```

```
Db     105839 AACAAAACATAATGTTGTAGCACTATACATTTTAAAAATTATTTTCTTTTGTTGGAACTA 105898

Qy        604 TTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATG 663
              ||  |  |  ||| |||       |   |  ||   ||   |||         | || | |
Db     105899 TTTAAATTTTAGGAATAAAATTCTAGCCACAAGACCTTAATAGGCACAGGATCAACGTTA 105958

Qy        664 TTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTC 723
              |  |   ||  ||  |   | |||      || |   ||||  |||||||||| ||||  ||
Db     105959 TCCTCAAACTT--ATTTTCCATTAGAAGAATTCCTGAAAACAGGTTGGGCACGGTGATTC 106016

Qy        724 ACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTT 783
              | ||||||| |||| ||||||||||||||| ||||    |  |||||||||| | ||||||
Db     106017 ATGCCTGTAACCCAACACTTTGGGAGGCTGAGGTGGGTCAATCACTTGAGGCCAGGAGTT 106076

Qy        784 TGAGACCAGCCTGGACAACATGATGAAACACCCTC-TCTAATAAAAATACAAAAATTAGC 842
              |||||||||||| |||| ||  |||||  |  || ||||  | ||||||||||||||||
Db     106077 CAAGACCAGCCTGGCCAACGTGGCGAAACTCTATCTTCTACTCAAAATACAAAAATTAGC 106136

Qy        843 TGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              ||||  ||||||||  || |||||  | ||||||||| |||   ||||||| ||| |||
Db     106137 TGGGAGTGGTGGCACACATCTGTAATACCAGCTACTCGGGAAGATGAGGCACGAGAATC 106195
```

RESULT 13
AC018868

LOCUS	AC018868 162100 bp DNA linear PRI 09-JAN-2002
DEFINITION	Homo sapiens BAC clone RP11-83J16 from 15, complete sequence.
ACCESSION	AC018868
VERSION	AC018868.4 GI:13443246
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 162100)
AUTHORS	Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
PUBMED	9847074
REFERENCE	2 (bases 1 to 162100)
AUTHORS	Grewal,N., Hawkins,M., Maupin,R., Gregory,S. and Scherger,E.
TITLE	The sequence of Homo sapiens BAC clone RP11-83J16
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 162100)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 162100)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (24-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 162100)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE	6 (bases 1 to 162100)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Mar 24, 2001 this sequence version replaced gi:7630827. ---------- Genome Center Center: Washington University Genome Sequencing Center Center code: WUGSC Web site: http://genome.wustl.edu/gsc Contact: sapiens@watson.wustl.edu ---------- Summary Statistics Center project name: H_NH0083J16 ----------. NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions. This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest. MAPPING INFORMATION: Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc SOURCE INFORMATION: The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) VECTOR: pBACe3.6 NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is RP11-814P5. Actual start of this clone is at base position 1 of RP11-83J16; actual end is at base position 162100 of RP11-83J16. The sequence from base position 136368 to 136447 was derived from PCR product of RP11-83J16 BAC DNA.

FEATURES	Location/Qualifiers
source	1. .162100 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="15" /map="15" /clone="RP11-83J16" /clone_lib="RPCI-11"
repeat_region	31. .332 /rpt_family="L1"
repeat_region	268. .309 /rpt_family="(GA)n"
repeat_region	852. .1141 /rpt_family="L1"
repeat_region	1142. .1430 /rpt_family="Alu"
repeat_region	1431. .1594 /rpt_family="L1"
repeat_region	1886. .3527 /rpt_family="L1"
repeat_region	2849. .2877 /rpt_family="AT_rich"
repeat_region	3598. .4454 /rpt_family="L1"
repeat_region	4595. .4618 /rpt_family="(A)n"
repeat_region	6431. .6622 /rpt_family="L2"
repeat_region	6923. .7015 /rpt_family="L2"
misc_feature	7129. .7149 /note="similar to Homo sapiens EST BE560643 (NID:g9804363)"

```
repeat_region   7131. .7356
                /rpt_family="Alu"
repeat_region   7295. .7322
                /rpt_family="AT_rich"
repeat_region   7473. .7582
                /rpt_family="MER1_type"
misc_feature    7504. .7992
                /note="similar to Sus scrofa EST BG381987 (NID:g13306459)"
repeat_region   8013. .8164
                /rpt_family="MER1_type"
repeat_region   8678. .8882
                /rpt_family="MIR"
repeat_region   8912. .9211
                /rpt_family="Alu"
repeat_region   9031. .9061
                /rpt_family="AT_rich"
repeat_region   9195. .9215
                /rpt_family="AT_rich"
repeat_region   9212. .9312
                /rpt_family="MIR"
repeat_region   9653. .9695
                /rpt_family="(CTATT)n"
repeat_region   9768. .9791
                /rpt_family="AT_rich"
repeat_region   9833. .10006
                /rpt_family="MER1_type"
repeat_region   10012. .10133
                /rpt_family="L2"
misc_feature    10065. .10079
                /note="similar to Homo sapiens EST BE278798
                (NID:g9153784)"
repeat_region   10275. .10369
                /rpt_family="L2"
repeat_region   11854. .11931
                /rpt_family="ERV1"
repeat_region   11926. .12288
                /rpt_family="ERV1"
repeat_region   13262. .13418
                /rpt_family="Alu"
repeat_region   13735. .13915
                /rpt_family="L1"
repeat_region   13893. .13924
                /rpt_family="AT_rich"
misc_feature    14140. .14156
                /note="similar to Homo sapiens EST AA808895 (NID:g2878301)
                nw16b06.s1"
repeat_region   14195. .14217
                /rpt_family="AT_rich"
repeat_region   14646. .14709
                /rpt_family="CT-rich"
repeat_region   14721. .14758
                /rpt_family="(CAT)n"
repeat_region   15600. .15775
                /rpt_family="L1"
repeat_region   15884. .16117
                /rpt_family="L1"
repeat_region   16118. .16241
                /rpt_family="Alu"
repeat_region   16242. .16907
                /rpt_family="L1"
repeat_region   16547. .16574
                /rpt_family="AT_rich"
repeat_region   16908. .17191
                /rpt_family="Alu"
repeat_region   17192. .17844
                /rpt_family="L1"
repeat_region   17856. .17975
                /rpt_family="Alu"
repeat_region   18010. .18814
                /rpt_family="L1"
repeat_region   18928. .19231
                /rpt_family="Alu"
repeat_region   19210. .19231
                /rpt_family="(A)n"
repeat_region   19412. .19772
                /rpt_family="MaLR"
repeat_region   19960. .21700
                /rpt_family="L1"
repeat_region   21702. .22516
```

Query Match 36.0%; Score 324.2; DB 8; Length 162100;
Best Local Similarity 62.7%; Pred. No. 1.2e-55;
Matches 564; Conservative 0; Mismatches 313; Indels 22; Gaps 3;

```
Qy          4 TCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 63
              ||| ||     || ||    |     |||  ||||| |||| |||| ||| | ||| |||
Db     105325 TCAAATATGAAAATAAGTGTTACACCGGGTGCAGTGGCTCATGCCTTTAATCCCAGCACT 105384

Qy         64 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 123
              |||| || |  | |  |||||||||| |||||||| ||||| |||| |||||||  |||
Db     105385 TTGGGAGGCTGAAGCAGGTGGATCACCTGAGGTGAGGAGTTCAAGACTAGCCTGACCAAC 105444

Qy        124 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 183
              |||||||||||| ||||||||||||||||||||||||||||||| || ||| ||||  ||
Db     105445 ATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGCACACC 105504

Qy        184 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 243
              |||| ||||||||||||| |||||||||||||||||||||||||||||| ||||||| ||
Db     105505 TGTAATCCCAGCTACTCAGGAGGCTGAGGCAAGAGAATTGCTTGAACCCGGGAGGCGGAA 105564

Qy        244 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 303
              |||||||||| |||||||  ||||| |||||||||||||||| |||| || |||||||  |
Db     105565 GTTGCAGTGAACCGAGATTGCACCATTGCACTCCAGCCTGGGGGACAGAGTGAGACTCCA 105624

Qy        304 TCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCC 363
              |||||||||| ||| |||  ||| |||       | |       | |||||||
Db     105625 TCTCAAAAAAAAAAAAAAAAGAAAGAAAAAGAAAAATAAGTGTTTATCTAATAA------ 105678

Qy        364 TGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTA 423
                        | ||     |   | | | || || | |     | || | || |
Db     105679 ------------TGCTAAAATGACATATTCTTAAAATGTATCACATTCTCTTCAGTGAA 105725

Qy        424 ATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACT 483
                  |  ||||       || |   ||   || | | | |||
Db     105726 GATAGTATAATTATAGAAAAAATCATGATATCTAAACAATATGTATGTTTTAACTTTTTG 105785

Qy        484 CTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAA 543
              |  || |   || ||    ||    ||| |      ||||||       |  ||  |
Db     105786 ATCATAAATTGAATAGTAGAAGAATCCAAAGCCTGGGGAAATCCTTTAGGGAAGGGAGGA 105845

Qy        544 CCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAAT 603
                |||||  |  |  ||   | |      || |     |||| || | | ||||
Db     105846 AACAAAACATAATGTTGTAGCACTATACATTTTAAAAATTATTTTCTTTTGTTGGAACTA 105905

Qy        604 TTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATG 663
              ||  |  |  ||| |||       |   |  ||   ||  |||        | || | |
Db     105906 TTTAAATTTTAGGAATAAAATTCTAGCCACAAGACCTTAATAGGCACAGGATCAACGTTA 105965

Qy        664 TTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTC 723
              | |   || ||  |   | |||       || |   ||||  |||||||||| ||||  ||
Db     105966 TCCTCAAACTT--ATTTTCCATTAGAAGAATTCCTGAAAACAGGTTGGGCACGGTGATTC 106023

Qy        724 ACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTT 783
              | ||||||| |||| ||||||||||||| ||||    |  ||||||||||| | ||||||
Db     106024 ATGCCTGTAACCCAACACTTTGGGAGGCTGAGGTGGGTCAATCACTTGAGGCCAGGAGTT 106083

Qy        784 TGAGACCAGCCTGGACAACATGATGAAACACCCTC-TCTAATAAAAATACAAAAATTAGC 842
              ||||||||||||  |||| ||  ||||| |  || |||| | ||||||||||||||||
Db     106084 CAAGACCAGCCTGGCCAACGTGGCGAAACTCTATCTTCTACTCAAAATACAAAAATTAGC 106143

Qy        843 TGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              ||||  ||||||||  || ||||| | ||||||||||  |||   ||||||| ||| |||
Db     106144 TGGGAGTGGTGGCACACATCTGTAATACCAGCTACTCGGGAAGATGAGGCACGAGAATC 106202

RESULT 14
```

```
AC090783/c
LOCUS       AC090783              180230 bp    DNA     linear   HTG 30-JUL-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-205B5 map 15, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC090783
VERSION     AC090783.2  GI:15028544
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 180230)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-205B5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180230)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 29, 2001 this sequence version replaced gi:13270652.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12756
                Center clone name: 205_B_5
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 176672 bases at least Q40
                Consensus quality: 177894 bases at least Q30
                Consensus quality: 178437 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 179230; sum-of-contigs
                Quality coverage: 6.8 in Q20 bases; agarose-fp
                Quality coverage: 6.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    63929: contig of 63929 bp in length
            *   63930    64029: gap of 100 bp
            *   64030    66793: contig of 2764 bp in length
            *   66794    66893: gap of 100 bp
            *   66894    69526: contig of 2633 bp in length
            *   69527    69626: gap of 100 bp
            *   69627    72423: contig of 2797 bp in length
            *   72424    72523: gap of 100 bp
            *   72524    76915: contig of 4392 bp in length
            *   76916    77015: gap of 100 bp
            *   77016    81564: contig of 4549 bp in length
            *   81565    81664: gap of 100 bp
            *   81665    94020: contig of 12356 bp in length
            *   94021    94120: gap of 100 bp
            *   94121   107816: contig of 13696 bp in length
            *  107817   107916: gap of 100 bp
            *  107917   136957: contig of 29041 bp in length
            *  136958   137057: gap of 100 bp
            *  137058   168034: contig of 30977 bp in length
            *  168035   168134: gap of 100 bp
            *  168135   180230: contig of 12096 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180230
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-205B5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .63929
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             63930. .64029
                     /estimated_length=100
     misc_feature    64030. .66793
                     /note="assembly_fragment"
     gap             66794. .66893
                     /estimated_length=100
     misc_feature    66894. .69526
                     /note="assembly_fragment"
     gap             69527. .69626
                     /estimated_length=100
     misc_feature    69627. .72423
                     /note="assembly_fragment"
     gap             72424. .72523
                     /estimated_length=100
     misc_feature    72524. .76915
                     /note="assembly_fragment"
     gap             76916. .77015
                     /estimated_length=100
     misc_feature    77016. .81564
                     /note="assembly_fragment"
     gap             81565. .81664
                     /estimated_length=100
     misc_feature    81665. .94020
                     /note="assembly_fragment"
     gap             94021. .94120
                     /estimated_length=100
     misc_feature    94121. .107816
                     /note="assembly_fragment"
     gap             107817. .107916
                     /estimated_length=100
     misc_feature    107917. .136957
                     /note="assembly_fragment"
     gap             136958. .137057
                     /estimated_length=100
     misc_feature    137058. .168034
                     /note="assembly_fragment"
     gap             168035. .168134
                     /estimated_length=100
     misc_feature    168135. .180230
```

```
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             36.0%;  Score 324.2;  DB 14;  Length 180230;
  Best Local Similarity   62.7%;  Pred. No. 1.2e-55;
  Matches  564;  Conservative    0;  Mismatches  313;  Indels   22;  Gaps    3;

Qy          4 TCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 63
              ||| ||     || ||    |     |||  ||||| |||| |||| ||| | ||| |||
Db     164565 TCAAATATGAAAATAAGTGTTACACCGGGTGCAGTGGCTCATGCCTTTAATCCCAGCACT 164506

Qy         64 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 123
              |||| || |  | |  |||||||||| |||||||| ||||| |||| ||||||  |||
Db     164505 TTGGGAGGCTGAAGCAGGTGGATCACCTGAGGTGAGGAGTTCAAGACTAGCCTGACCAAC 164446

Qy        124 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 183
              |||||||||||| |||||||||||||||||||||||||||||||| || ||| ||||  ||
Db     164445 ATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGCACACC 164386

Qy        184 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 243
              |||| ||||||||||||| |||||||||||||||||||||||||||| ||||||| ||
Db     164385 TGTAATCCCAGCTACTCAGGAGGCTGAGGCAAGAGAATTGCTTGAACCCGGGAGGCGGAA 164326

Qy        244 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 303
              |||||||||| |||||||  ||||| ||||||||||||||| |||| || ||||||| |
Db     164325 GTTGCAGTGAACCGAGATTGCACCATTGCACTCCAGCCTGGGGGACAGAGTGAGACTCCA 164266

Qy        304 TCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCC 363
              |||||||||||  ||| |||  ||| |||       | |       | |||||||
Db     164265 TCTCAAAAAAAAAAAAAAAAGAAAGAAAAAGAAAAATAAGTGTTTATCTAATAA------ 164212

Qy        364 TGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTA 423
                        | ||     |     | | | || || | |     | || | || |
Db     164211 -------------TGCTAAAATGACATATTCTTAAAATGTATCACATTCTCTTCAGTGAA 164165

Qy        424 ATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACT 483
                  |  ||||       || |   ||   || | | | |||
Db     164164 GATAGTATAATTATAGAAAAAATCATGATATCTAAACAATATGTATGTTTTAACTTTTTG 164105

Qy        484 CTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAA 543
                |  || |   || ||    ||    ||| |       ||||||       |  ||  |
Db     164104 ATCATAAATTGAATAGTAGAAGAATCCAAAGCCTGGGGAAATCCTTTAGGGAAGGGAGGA 164045

Qy        544 CCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAAT 603
                |||||  |   |  ||   | |      || |     |||| || | | ||||
Db     164044 AACAAAACATAATGTTGTAGCACTATACATTTTAAAAATTATTTTCTTTTGTTGGAACTA 163985

Qy        604 TTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATG 663
              ||  |  |  ||| |||        |   |  ||   ||    |||         | || | |
Db     163984 TTTAAATTTTAGGAATAAAATTCTAGCCACAAGACCTTAATAGGCACAGGATCAACGTTA 163925

Qy        664 TTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTC 723
              | |   || ||   |   | |||      || |   ||||  |||||||||| ||||  ||
Db     163924 TCCTCAAACTT--ATTTTCCATTAGAAGAATTCCTGAAAACAGGTTGGGCACGGTGATTC 163867

Qy        724 ACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTT 783
              | ||||||| |||| |||||||||||||| ||||    |  ||||||||||| | ||||||
Db     163866 ATGCCTGTAACCCAACACTTTGGGAGGCTGAGGTGGGTCAATCACTTGAGGCCAGGAGTT 163807

Qy        784 TGAGACCAGCCTGGACAACATGATGAAACACCCTC-TCTAATAAAAATACAAAAATTAGC 842
                |||||||||||| |||| ||  ||||| |  || |||| | ||||||||||||||||
Db     163806 CAAGACCAGCCTGGCCAACGTGGCGAAACTCTATCTTCTACTCAAAATACAAAAATTAGC 163747

Qy        843 TGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              ||||  ||||||||  || ||||| | ||||||||||| |||   ||||||| ||| |||
Db     163746 TGGGAGTGGTGGCACACATCTGTAATACCAGCTACTCGGGAAGATGAGGCACGAGAATC 163688


RESULT 15
AL591215/c
LOCUS       AL591215              154592 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-20D12 on chromosome 9 Contains
            the 3' end of a novel gene, possible ortholog of mouse RIKEN cDNA
            1700028P14 gene, complete sequence.
ACCESSION   AL591215 AC025064
VERSION     AL591215.12  GI:15808225
KEYWORDS    HTG; 1700028P14.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 154592)
  AUTHORS   Garner,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 28, 2001 this sequence version replaced gi:15723808.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-20D12 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .154592
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-20D12"
                     /clone_lib="RPCI-11.1"
     misc_feature    1
                     /note="Clone_left_end: RP11-20D12"
     gene            join(AL162412.11:177963. .178201,3018. .3126,15056. .15174,
                     16410. .16478,45339. .45396,64455. .64733)
                     /locus_tag="RP11-20D12.1-001"
     mRNA            join(AL162412.11:177963. .178201,3018. .3126,15056. .15174,
                     16410. .16478,45339. .45396,64455. .64733)
                     /locus_tag="RP11-20D12.1-001"
                     /note="match: ESTs: AA416734.1 AA416815.1 AA757101.1
                     AI768674.1 AW654647.1 BB573244.1 BF059207.1 BG193301.1
                     BG194901.1 BU944256.1 BU961648.1
                     match: cDNAs: AK006473.1 AK014972.1"
     CDS             join(AL162412.11:178050. .178201,3018. .3126,15056. .15174,
                     16410. .16478,45339. .45396,64455. .64637)
                     /locus_tag="RP11-20D12.1-001"
                     /standard_name="OTTHUMP00000021439"
                     /note="match: proteins: Q9CQC3"
```

```
                     /codon_start=1
                     /protein_id="CAH71316.1"
                     /db_xref="GI:55665830"
                     /db_xref="UniProt/TrEMBL:Q5VTT2"
                     /translation="MDSLDRSCQDWCDRKQHWLEIGPPDLVERKGSLTLRSHHKKYSK
                     PVLVYSWHRDREAFPKGYDIEGPEKVKKLCNSTYRRLGTDESPIWTSETHEKLSQMCL
                     NTEWVEMKSKALLNEETVSSGIIERVTGLPATGFGAVFPRHPPDWSKMCALTTYSEDY
                     VPPYDYQPHAYPCQDDYSIVHRKCRSQFTDLNGSKRFGINTWHDESGIYANSDVKQKL
                     YPLTSGPIVPI"
     gene            join(AL162412.11:178130. .178201,3018. .3126,15056. .15174,
                     16410. .16659)
                     /locus_tag="RP11-20D12.1-002"
     mRNA            join(AL162412.11:178130. .178201,3018. .3126,15056. .15174,
                     16410. .16659)
                     /locus_tag="RP11-20D12.1-002"
                     /note="match: ESTs: AA431486.1 AA948291.1 AW014133.1
                     BE502978.1 BG192998.1 BI825449.1"
     gene            join(AL162412.11:178090. .178201,15056. .15174,16410. .16680)
                     /locus_tag="RP11-20D12.1-003"
     mRNA            join(AL162412.11:178090. .178201,15056. .15174,16410. .16680)
                     /locus_tag="RP11-20D12.1-003"
                     /note="match: ESTs: AW014133.1"
     polyA_signal    16661. .16666
                     /locus_tag="RP11-20D12.1-003"
     polyA_site      16680
                     /locus_tag="RP11-20D12.1-003"
     gene            join(43556. .43623,45339. .45396,64455. .64728)
                     /locus_tag="RP11-20D12.1-004"
     mRNA            join(43556. .43623,45339. .45396,64455. .64728)
                     /locus_tag="RP11-20D12.1-004"
                     /note="match: ESTs: BG194901.1"
     polyA_signal    64707. .64712
                     /locus_tag="RP11-20D12.1-004"
     polyA_site      64728
                     /locus_tag="RP11-20D12.1-004"
     misc_feature    126192
                     /note="Clone_right_end: RP11-385E2"
     misc_feature    129644
                     /note="Clone_left_end: RP11-44J18"
     misc_feature    154592
                     /note="Clone_right_end: RP11-20D12"
ORIGIN

  Query Match             35.9%;  Score 323.8;  DB 8;  Length 154592;
  Best Local Similarity   62.2%;  Pred. No. 1.4e-55;
  Matches  564;  Conservative    0;  Mismatches  332;  Indels   11;  Gaps    3;

Qy          4 TCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 63
              | | || |    |||  |   | | ||||||||||| |||| |||||||| | ||| |||
Db     117733 TAAAATAAGAGTAAATAATACCGGCTGGGCACAGTGGCTCATGCCTGTAATCCCAGCACT 117674

Qy         64 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 123
              |||| ||    ||| |||  ||||| ||||||  |  ||||| |||||||||||| |||
Db     117673 TTGGGAGGTGGAGGCGGGCAGATCATTTGAGGACAAGAGTTCAAGACCAGCCTGGGCAAC 117614

Qy        124 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 183
              |||  |||||||| |||||||| ||||||||||||||||||||||| ||| ||||||||
Db     117613 ATGACGAAACCCTGTCTCTACTGAAAATACAAAAATTAGCTGGGTGTGGTGGTGCATGCC 117554

Qy        184 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 243
              |||| |||||||| || |||||||||||||  |||||  |||||||||||||| | |||
Db     117553 TGTAATCCCAGCTGCTTGGGAGGCTGAGGCACAAGAATCACTTGAACCTGGGAGACGGAG 117494

Qy        244 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 303
              |||||||||||| ||||||  |||||||||||  |||||||||||||| ||| |||||||
Db     117493 GTTGCAGTGAGCTGAGATCTTACCACTGCACTACAGCCTGGGCGACAGAGCAAGACTCTG 117434

Qy        304 TCTCAAAAAAAATAAATAAATAAAATAAA---------GGATCGGAGAGAAACAAAACTAA 354
              ||||||||||  ||| |  ||||| |||          | || | |   |       |  |
Db     117433 TCTCAAAAAATAAAAAAGTTAAAAAAAAAAGAAGAGTGAATAGTACCTACCACGGAAATA 117374

Qy        355 TAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATT 414
              |   ||      | | || ||||| ||| |     || | || |  |        || |

Db     117373 TTTCATGAAGATAAGAAAACAGAGGTACTTGGTACATGTCTACTGCATAGAGTGAGCTTA 117314

Qy        415 TTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAA 474
              ||| | | | ||  |  |    |  || || ||||||  |    |   | | |
Db     117313 GTAATTCTTACCTATTCATAATAATAACTTCTAGAAGTAATTATAAAATCTAAATTATGT 117254

Qy        475 CTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACC 534
              |  |   ||| ||  || | ||  |   || | |       | || ||| || | |
Db     117253 TTAAATTATCTTCAACTATCTTTAGAAATCAATG-GCTGTTGAACTGTAACCCAATACTT 117195

Qy        535 CAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATT 594
              | | | |    |  |  |||       |             |   |||         | |
Db     117194 TATACGTCATTCTTCCAAGGTTCAGGACAACCCCAAAATATGATGGTAGGAGACCAGAAT 117135

Qy        595 TGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCAT 654
              |  |    ||||    | |  |  || |||| |    ||   || |   |
Db     117134 ATGCCAGTCCAAAATGTAATCACCTGACAGGTTGTTCTTACCTGCTGCCCAGAAAAGCCA 117075

Qy        655 TTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCA 714
                ||                | ||      ||| |    ||    ||   |||
Db     117074 AATGCACTGAGAACAGCAGGAGTGTTTGCAGCAAAGAAAATGTTTAATTGTGGCCAGGCG 117015

Qy        715 TGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAG 773
              |||||||||| |||||| |||||||||||||||||||| |||||| |||||||||| ||
Db     117014 TGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGATGGATCACTTAAG 116955

Qy        774 GTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACA 833
              ||| |||||||||||||||||| ||||||| |||||| || | |||| |||||||| |
Db     116954 GTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTATCTACTAAAAATATA 116895

Qy        834 AAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCA 893
              |||||||||    | ||||||  ||  |||||| |||||||||||||||||| |||||| |
Db     116894 AAAATTAGCCTAGTGTGGTGGTGAGTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGTA 116835

Qy        894 GGAGTAT 900
              |||  ||
Db     116834 GGAAAAT 116828


Search completed: February  9, 2006, 20:42:30
Job time : 4495.31 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 580.391 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 901
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	901	100.0	12174	12	ADJ09983	Adj09983 Human pho
	2	900.6	100.0	12174	12	ADI35082	Adi35082 Human PLA
	3	899.4	99.8	13612	6	ABK47376	Abk47376 Human Pho
c	4	555.4	61.6	42360	13	ABD33466	Abd33466 Human can
	5	313.6	34.8	22976	2	AAX83426	Aax83426 Genomic r
	6	313.6	34.8	23187	3	AAA50273	Aaa50273 Human lip
	7	313.6	34.8	23187	4	AAF62331	Aaf62331 Human lep
	8	313.6	34.8	41936	6	ABL67924	Abl67924 Ovary can
	9	313.6	34.8	56580	13	ADT25932	Adt25932 Wild type
	10	313.6	34.8	56580	13	ADU23759	Adu23759 Human ATP
	11	312.6	34.7	21721	2	AAX83427	Aax83427 Human lip
c	12	312.4	34.7	37940	13	ADS19371	Ads19371 Human PIP
c	13	312.4	34.7	38000	12	ADJ26093	Adj26093 Human PIP
c	14	310	34.4	28818	4	AAL35901	Aal35901 Human mus
c	15	310	34.4	28818	8	ABX58889	Abx58889 cDNA enco
c	16	310	34.4	28818	12	ADJ29639	Adj29639 Human mus
	17	308.2	34.2	196686	11	ACN44170	Acn44170 Human gen
	18	307.2	34.1	69770	10	ADC86870	Adc86870 Human GPC
c	19	307	34.1	234882	14	ADZ13715	Adz13715 Human can
c	20	307	34.1	246940	12	ADQ59422	Adq59422 Human can
c	21	305.2	33.9	23434	8	ACC00503	Acc00503 Human his
c	22	304	33.7	260160	12	ADQ20017	Adq20017 Human sof
	23	303.8	33.7	32250	4	AAL05627	Aal05627 Human rep
	24	303.2	33.7	114693	8	AAD48308	Aad48308 Human tra
c	25	302.8	33.6	14598	4	AAK80690	Aak80690 Human imm
c	26	302.8	33.6	14598	4	AAK79627	Aak79627 Human imm
c	27	302.6	33.6	174472	14	ADZ13139	Adz13139 Human can
c	28	302.6	33.6	174703	11	ACN44738	Acn44738 Human gen
	29	302.2	33.5	38918	10	ADC87242	Adc87242 Human GPC
	30	301.8	33.5	2611	6	ABS76421	Abs76421 cDNA enco
	31	301.8	33.5	2611	14	ADZ51348	Adz51348 Nucleotid
c	32	301.6	33.5	59588	10	ADC87426	Adc87426 Human GPC
c	33	301.6	33.5	301477	13	ABD33362	Abd33362 Human can
c	34	300.4	33.3	167163	10	ADE82948	Ade82948 Human PVT
c	35	300.2	33.3	37160	6	ABK84372	Abk84372 Human cDN
	36	299.4	33.2	1341	6	ABK14184	Abk14184 Human spl
c	37	297.4	33.0	110000	11	ACN44014_2	Continuation (3 of
	38	297.2	33.0	16869	3	ABN97976	Abn97976 Human ret
c	39	297.2	33.0	56093	6	ABL61744	Abl61744 Colon ade
c	40	297.2	33.0	119501	12	ADI29095	Adi29095 Human MAR
	41	297	33.0	133632	11	ACN45054	Acn45054 Human gen
	42	297	33.0	133642	14	ADZ13285	Adz13285 Human can
	43	296.4	32.9	110000	14	AEA61163_1	Continuation (2 of
	44	295.8	32.8	103747	6	ABQ88139	Abq88139 Human ost
	45	295.2	32.8	8458	4	AAK72176	Aak72176 Human imm

ALIGNMENTS

```
RESULT 1
ADJ09983
ID   ADJ09983 standard; DNA; 12174 BP.
XX
AC   ADJ09983;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW   human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW   PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW   appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW   non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW   hypertension; antidiabetic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(436,c)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(839,a)
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4050,a)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4689,t)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6282,a)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6358,c)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6653,t)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7256,t)
FT                   /*tag= h
```

```
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7300,a)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7301,c)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7328,g)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(8062,c)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(9182,t)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(11649,c)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphisms"
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   P-PSDB; ADJ09984.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ   Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 901;  DB 12;  Length 12174;
  Best Local Similarity   100.0%;  Pred. No. 3e-187;
  Matches  901;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8700 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8759

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8760 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8819

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8820 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8879

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8880 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8939

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8940 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8999

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9000 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 9059

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9060 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 9119

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9120 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 9179

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9180 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 9239

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9240 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 9299

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9300 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 9359

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9360 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 9419

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9420 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 9479

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9480 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 9539

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9540 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 9599

Qy        901 C 901
              |
Db       9600 C 9600


RESULT 2
ADI35082
ID   ADI35082 standard; DNA; 12174 BP.
XX
AC   ADI35082;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human PLA2G1B nucleotide sequence.
XX
KW   PLA2G1B ; fat deposition; leanness; polymorphism;
KW   non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW   hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW   microalbuminuria; human; gene; ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO2004002295-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020830.
XX
PR   27-JUN-2002; 2002US-0392361P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML;
XX
DR   WPI; 2004-082843/08.
DR   P-PSDB; ADI35083.
XX
PT   Diagnosing a predisposition to fat deposition or leanness, useful for
PT   diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT   detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT   the subject.
XX
PS   Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC   The invention relates to diagnosing a predisposition to fat deposition or
CC   leanness in a subject comprising detecting the presence or absence of a
CC   polymorphic variation associated with fat deposition at a polymorphic
CC   site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC   subject, where the presence of the polymorphic variation indicates a
CC   predisposition to fat deposition in the subject. The polymorphic
CC   variation is a guanine at position 7328 or thymine at position 9182 of
CC   the present sequence. The method is useful for diagnosing a
CC   predisposition to fat deposition or leanness in a subject, and
CC   consequently for diagnosing a predisposition to non-insulin dependent
CC   diabetes mellitus (NIDDM) in a subject and conditions such as
CC   hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC   hypercoagulability, or microalbuminuria, which can lead to early
CC   prescription of preventive measures. The present sequence represents a
CC   human PLA2G1B nucleotide sequence.
XX
SQ   Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 901;  DB 12;  Length 12174;
  Best Local Similarity   99.9%;  Pred. No. 3.7e-187;
  Matches  900;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8700 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8759

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8760 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8819

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8820 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8879

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8880 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8939

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8940 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8999

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9000 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 9059

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9060 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 9119

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9120 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 9179

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9180 ACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 9239

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9240 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 9299

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9300 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 9359

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9360 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 9419

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9420 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 9479

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9480 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 9539

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9540 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 9599

Qy        901 C 901
              |
Db       9600 C 9600


RESULT 3
ABK47376
ID   ABK47376 standard; DNA; 13612 BP.
XX
AC   ABK47376;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW   Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(3845,A)
FT                   /*tag=  a
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(3968,A)
FT                   /*tag=  b
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4053. .9613
FT                   /*tag=  c
FT                   /product= "Pla2G1B"
FT   exon            4053. .4086
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4087. .5785
FT                   /*tag=  e
FT                   /number= 1
FT   exon            5786. .5945
FT                   /*tag=  f
```

```
FT                   /number= 2
FT   intron          5946. .6744
FT                   /*tag=  g
FT                   /number= 2
FT   variation       replace(6060,A)
FT                   /*tag=  h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   exon            6745. .6873
FT                   /*tag=  i
FT                   /number= 3
FT   variation       replace(6844,A)
FT                   /*tag=  j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   intron          6874. .9489
FT                   /*tag=  k
FT                   /number= 3
FT   exon            9490. .9613
FT                   /*tag=  l
FT                   /number= 4
FT   variation       replace(9531,A)
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             99.8%;  Score 899.4;  DB 6;  Length 13612;
  Best Local Similarity   99.9%;  Pred. No. 6.9e-187;
  Matches  900;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8275

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8335

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8395

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8455

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8515

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 8575

Qy        361 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 8635

Qy        421 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 8695

Qy        481 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8696 ACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGA 8755

Qy        541 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8756 TAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAA 8815

Qy        601 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8816 AATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGG 8875

Qy        661 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8876 ATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGG 8935

Qy        721 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8936 CTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 8995

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8996 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTA 9055

Qy        841 GCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       9056 GCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 9115

Qy        901 C 901
              |
Db       9116 C 9116
```

```
RESULT 4
ABD33466/c
ID   ABD33466 standard; DNA; 42360 BP.
XX
AC   ABD33466;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-089.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 618; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 42360 BP; 11698 A; 9658 C; 10445 G; 10559 T; 0 U; 0 Other;

  Query Match             61.6%;  Score 555.4;  DB 13;  Length 42360;
  Best Local Similarity   99.8%;  Pred. No. 1.6e-111;
  Matches  556;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        345 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42360 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 42301

Qy        405 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42300 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 42241

Qy        465 AGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42240 AGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 42181

Qy        525 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42180 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 42121

Qy        585 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42120 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 42061

Qy        645 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42060 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 42001

Qy        705 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42000 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 41941

Qy        765 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAAT 824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41940 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAAT 41881

Qy        825 AAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAG 884
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      41880 AAAAATACAAAAATTAGCTGGGCATGGTGGCGAGCACCTGTAGTCCCAGCTACTCAGGAG 41821

Qy        885 TCTGAGGCAGGAGTATC 901
              |||||||||||||||||
Db      41820 TCTGAGGCAGGAGTATC 41804


RESULT 5
AAX83426
ID   AAX83426 standard; DNA; 22976 BP.
XX
AC   AAX83426;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Genomic region containing human lipolysis stimulated receptor gene.
XX
KW   Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW   cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW   allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW   transgenic animal; gene expression; triglyceride; eating disorder;
KW   obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW   anorexia; metabolism; ds.
XX
OS   Homo sapiens.
XX
PN   WO9907737-A2.
XX
PD   18-FEB-1999.
XX
PF   06-AUG-1998;   98WO-IB001257.
XX
PR   06-AUG-1997;   97FR-00010088.
PR   22-APR-1998;   98FR-00005032.
XX
PA   (GEST ) GENSET.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bihain B,  Bougueleret L,  Yen-Potin F;
XX
DR   WPI; 1999-190035/16.
XX
PT   Lipolysis stimulated receptor involved in leptin metabolism - and
PT   controlling its activity for treatment of obesity, anorexia or abnormal
PT   cytokine metabolism.
XX
PS   Claim 40; Page 222-242; 279pp; French.
XX
CC   This sequence represents the genomic region containing the coding region
CC   for the human lipolysis stimulated receptor (LSR). The novel LSR binds
CC   lipoproteins in presence of free fatty acids (FFA) and binds cytokines in
CC   absence of FFA. The nucleic acid and its fragments are used as probes and
CC   primers for detection and/or amplification of LSR genes; for production
```

```
CC   of recombinant LSR; for detecting allelic variants, mutations, deletions,
CC   loss of heterozygosity and genetic abnormalities in the gene. LSR,
CC   recombinant cells and transgenic animals are used to screen for chemical
CC   interacting with LSR, also to study expression and activity of LSR and
CC   its interactions. The chemicals, and leptin, are used to modulate the
CC   number of LSR in a cell, its recycling rate and/or specificity of
CC   receptor activity, particularly for reducing the level of leptin,
CC   lipoproteins, chylomicrons and/or triglycerides. The chemicals are thus
CC   useful for treating eating disorders, particularly obesity (and related
CC   diseases such as atheromatosis, atherosclerosis, hypertension and
CC   diabetes) or anorexia, also disease associated with abnormal cytokine
CC   metabolism
XX
SQ   Sequence 22976 BP; 5120 A; 6191 C; 6677 G; 4988 T; 0 U; 0 Other;

  Query Match             34.8%;  Score 313.6;  DB 2;  Length 22976;
  Best Local Similarity   62.4%;  Pred. No. 1.2e-58;
  Matches  545;  Conservative    0;  Mismatches  319;  Indels   10;  Gaps    3;

Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88
Db       5491 TGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA 5550

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
Db       5551 CCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAA 5610

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
Db       5611 AATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 5670

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
Db       5671 GAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCACCA 5730

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
Db       5731 CTGTACTGCAGTCTGGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAA 5790

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
Db       5791 GACTCCGTCAAGGTATAAGAATGTCAGAGAGTACTAAGTGTTGCAAAG--------AAAA 5842

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
Db       5843 TAACACCAGGCTGGGTGCATTGGCTCATGCCTGTAAATTTCAGCACTTTGGGAGGCCAAG 5902

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
Db       5903 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACAACAAAATGAGACCCC 5962

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
Db       5963 ATGTCTACAAAAATTTTAAAAATTTAAAAATTAGCTGGGCATGGTGGCATGTGCCTGTGG 6022

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
Db       6023 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 6082

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
Db       6083 AGTGAGTCATGATCGTGCCACTGCATTCCAGCCTGGGTGA-CAGAGTGAGACCCTGTCTT 6141

Qy        689 TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGG 747
Db       6142 GAAATGAAAAGAAAATAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 6201

Qy        748 AGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGAT 807
Db       6202 AGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCAGCCTGGCCAACATGGT 6261

Qy        808 GAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867
Db       6262 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGGCGTGGTGGTGGGCGCCTGTAA 6321

Qy        868 TCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
Db       6322 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 6355


RESULT 6
AAA50273
ID   AAA50273 standard; DNA; 23187 BP.
XX
AC   AAA50273;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human lipolysis stimulated receptor (LSR) gene.
XX
KW   Lipolysis stimulated receptor; LSR; chromosome 19q13.1; human;
KW   single nucleotide polymorphism; biallelic marker; obesity;
KW   atherosclerosis; insulin resistance; hypertension; hyperlipidemia;
KW   hypertriglyceridemia; cardiovascular disease; microangiopathy;
KW   syndrome X; diagnosis; therapy; genotyping; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .2000
FT                   /*tag=  a
FT                   /note= "potential 5' regulatory region"
FT   primer_bind     complement(523. .544)
FT                   /*tag=  bj
FT                   /note= "upstream amplification primer 17-2"
FT   variation       replace(606,T)
FT                   /*tag=  am
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'1"
FT   variation       replace(818,G)
FT                   /*tag=  u
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 17-2-297 (A31)"
FT   primer_bind     complement(946. .963)
FT                   /*tag=  bl
FT                   /note= "upstream amplification primer 99-4576"
FT   primer_bind     1047. .1068
FT                   /*tag=  bk
FT                   /note= "downstream amplification primer 17-2"
FT   primer_bind     complement(1096. .1115)
FT                   /*tag=  bn
FT                   /note= "upstream amplification primer 9-19"
FT   variation       replace(1243,T)
FT                   /*tag=  v
FT                   /frequency= "0.15"
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-19-148 (A10)"
FT   variation       replace(1374,A)
FT                   /*tag=  w
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-19-256 (A32)"
FT   primer_bind     1385. .1402
FT                   /*tag=  bm
FT                   /note= "downstream amplification primer 99-4576"
FT   variation       replace(1401,T)
FT                   /*tag=  x
FT                   /frequency= "0.12"
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-19-307 (A11)"
FT   variation       replace(1534. .1535,C)
FT                   /*tag=  y
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-19-442 (A12)"
FT   primer_bind     complement(1602. .1621)
FT                   /*tag=  bp
FT                   /note= "upstream amplification primer 9-20"
```

```
FT   primer_bind     1616. .1635
FT                   /*tag= bo
FT                   /note= "downstream amplification primer 9-19"
FT   variation       replace(1788,A)
FT                   /*tag= z
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-20-187 (A13)"
FT   exon            2001. .2356
FT                   /*tag= b
FT                   /number= 1
FT   primer_bind     complement(2036. .2053)
FT                   /*tag= br
FT                   /note= "upstream amplification primer 99-4557"
FT   primer_bind     2062. .2081
FT                   /*tag= bw
FT                   /note= "downstream amplification primer 9-21"
FT   primer_bind     2074. .2093
FT                   /*tag= bq
FT                   /note= "downstream amplification primer 9-20"
FT   primer_bind     complement(2084. .2102)
FT                   /*tag= bt
FT                   /note= "upstream amplification primer 9-1"
FT   CDS             2104. .21082
FT                   /*tag= bi
FT                   /note= "contains introns, encodes AAY95822, AAY95823 and
FT                   AAY95824 as a result of alternative splicing"
FT   intron          2357. .3539
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2391,G)
FT                   /*tag= aa
FT                   /frequency= "0.24"
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-1-308 (A14)"
FT   primer_bind     complement(2470. .2489)
FT                   /*tag= bv
FT                   /note= "upstream amplification primer 9-21"
FT   primer_bind     2483. .2500
FT                   /*tag= bu
FT                   /note= "downstream amplification primer 9-1"
FT   primer_bind     2563. .2580
FT                   /*tag= bs
FT                   /note= "downstream amplification primer 99-4557"
FT   primer_bind     complement(3455. .3474)
FT                   /*tag= bx
FT                   /note= "upstream amplification primer 9-3"
FT   exon            3540. .3884
FT                   /*tag= d
FT                   /number= 2
FT   primer_bind     complement(3775. .3792)
FT                   /*tag= bz
FT                   /note= "upstream amplification primer 99-4558"
FT   variation       replace(3778,T)
FT                   /*tag= ab
FT                   /frequency= "0.29"
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 9-3-324 (A15)"
FT   primer_bind     3882. .3901
FT                   /*tag= by
FT                   /note= "downstream amplification primer 9-3"
FT   intron          3885. .12162
FT                   /*tag= e
FT                   /number= 2
FT   primer_bind     4336. .4356
FT                   /*tag= ca
FT                   /note= "downstream amplification primer 99-4558"
FT   primer_bind     4444. .44638
FT                   /*tag= cd
FT                   /note= "downstream amplification primer 99-14419"
FT   variation       replace(4498,T)
FT                   /*tag= ac
FT                   /frequency= "0.22"
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker 99-14419-424 (A16)"
FT   primer_bind     complement(4902. .4920)
FT                   /*tag= cb
FT                   /note= "upstream amplification primer 99-14419"
FT   variation       replace(5141,G)
FT                   /*tag= an
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'2"
FT   primer_bind     complement(6638. .6655)
FT                   /*tag= ce
FT                   /note= "upstream amplification primer 99-4577"
FT   primer_bind     7072. .7089
FT                   /*tag= cf
FT                   /note= "downstream amplification primer 99-4577"
FT   variation       replace(7428,C)
FT                   /*tag= ao
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'3"
FT   primer_bind     complement(7995. .8012)
FT                   /*tag= cg
FT                   /note= "upstream amplification primer 99-4559"
FT   variation       replace(8394,G)
FT                   /*tag= ap
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'4"
FT   primer_bind     8576. .8593
FT                   /*tag= ch
FT                   /note= "downstream amplification primer 99-4559"
FT   variation       replace(8704,C)
FT                   /*tag= aq
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'5"
FT   variation       replace(9028,A)
FT                   /*tag= ar
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'6"
FT   primer_bind     complement(9622. .9639)
FT                   /*tag= ci
FT                   /note= "upstream amplification primer 99-3148"
FT   variation       replace(9949. .9958,TA)
FT                   /*tag= as
FT                   /note= "marker A'7"
FT   primer_bind     complement(9964. .9981)
FT                   /*tag= ck
FT                   /note= "upstream amplification primer 99-4560"
FT   variation       replace(9977,C)
FT                   /*tag= at
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'8"
FT   variation       replace(10021,A)
FT                   /*tag= au
FT                   /standard_name= "single nucleotide polymorphism"
FT                   /note= "marker A'9"
FT   primer_bind     10023. .10040
FT                   /*tag= cj
FT                   /note= "downstream amplification primer 99-3148"
FT   primer_bind     10492. .10512
FT                   /*tag= cn
FT                   /note= "downstream amplification primer 99-14411"
FT   primer_bind     10546. .10563
FT                   /*tag= cl
FT                   /note= "downstream amplification primer 99-4560"
FT   primer_bind     complement(10996. .11015)
```

```
  Query Match             34.8%;  Score 313.6;  DB 3;  Length 23187;
  Best Local Similarity   62.4%;  Pred. No. 1.2e-58;
  Matches  545;  Conservative    0;  Mismatches  319;  Indels   10;  Gaps    3;

Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88
              ||||||| ||| |||| | |||||| | ||| |||||||| || ||||||||||||||||
Db       5595 TGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA 5654

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
```

```
Db       5655 CCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAA 5714

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208

Db       5715 AATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 5774

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268

Db       5775 GAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCACCA 5834

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328

Db       5835 CTGTACTGCAGTCTGGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 5894

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388

Db       5895 GACTCCGTCAAGGTATAAGAATGTCAGAGAGTACTAAGTGTTGCAAAG--------AAAA 5946

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448

Db       5947 TAACACCAGGCTGGGTGCATTGGCTCATGCCTGTAAATTTCAGCACTTTGGGAGGCCAAG 6006

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508

Db       6007 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACAACAAAATGAGACCCC 6066

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568

Db       6067 ATGTCTACAAAAATTTAAAAATTTAAAAATTAGCTGGGCATGGTGGCATGTGCCTGTGG 6126

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628

Db       6127 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 6186

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688

Db       6187 AGTGAGTCATGATCGTGCCACTGCATTCCAGCCTGGGTGA-CAGAGTGAGACCCTGTCTT 6245

Qy        689 TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGG 747

Db       6246 GAAATGAAAAGAAAATAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 6305

Qy        748 AGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGAT 807

Db       6306 AGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCAGCCTGGCCAACATGGT 6365

Qy        808 GAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867

Db       6366 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGGCGTGGTGGTGGGCGCCTGTAA 6425

Qy        868 TCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901

Db       6426 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 6459


RESULT 7
AAF62331
ID   AAF62331 standard; DNA; 23187 BP.
XX
AC   AAF62331;
XX
DT   06-JUN-2001  (first entry)
XX
DE   Human leptin fragment coding sequence SEQ ID NO: 1.
XX
KW   Leptin; human; LSR; lipolysis stimulated receptor; obesity; hypertension;
KW   anorexia; cachexia; stroke; atherosclerosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .2000
FT                   /*tag= a
FT                   /note= "5' regulatory region"
FT   primer_bind     523. .544
FT                   /*tag= b
FT   misc_binding    587. .605
FT                   /*tag= c
FT   allele          replace(606,T)
FT                   /*tag= d
FT   misc_binding    complement(607. .625)
FT                   /*tag= e
FT   misc_binding    799. .817
FT                   /*tag= f
FT   allele          replace(818,C)
FT                   /*tag= g
FT   misc_binding    complement(819. .837)
FT                   /*tag= h
FT   primer_bind     946. .963
FT                   /*tag= i
FT   primer_bind     complement(1047. .1068)
FT                   /*tag= j
FT   primer_bind     1096. .1115
FT                   /*tag= k
FT   misc_binding    1224. .1242
FT                   /*tag= l
FT   allele          replace(1243,T)
FT                   /*tag= m
FT   misc_binding    complement(1244. .1262)
FT                   /*tag= n
FT   misc_binding    1330. .1373
FT                   /*tag= o
FT   allele          replace(1374,G)
FT                   /*tag= p
FT   misc_binding    complement(1375. .1393)
FT                   /*tag= q
FT   misc_binding    1382. .1400
FT                   /*tag= r
FT   primer_bind     complement(1385. .1402)
FT                   /*tag= s
FT   allele          replace(1401,T)
FT                   /*tag= t
FT   misc_binding    complement(1402. .1420)
FT                   /*tag= u
FT   misc_binding    1516. .1534
FT                   /*tag= v
FT   primer_bind     1602. .1621
FT                   /*tag= w
FT   primer_bind     complement(1616. .1635)
FT                   /*tag= x
FT   misc_binding    1769. .1787
FT                   /*tag= y
FT   allele          replace(1788,C)
FT                   /*tag= z
FT   misc_binding    complement(1789. .1807)
FT                   /*tag= aa
FT   CDS             2001. .21190
FT                   /*tag= ab
FT                   /product= "LSR"
FT                   /note= "this sequence contains introns"
FT   exon            2001. .2356
FT                   /*tag= ac
FT                   /number= 1
FT   primer_bind     2036. .2053
FT                   /*tag= ad
FT   primer_bind     2062. .2081
FT                   /*tag= ae
FT   primer_bind     complement(2074. .2093)
FT                   /*tag= af
FT   primer_bind     2084. .2102
FT                   /*tag= ag
FT   misc_binding    2372. .2390
FT                   /*tag= ah
FT   allele          replace(2391,C)
FT                   /*tag= ai
```

```
FT   misc_binding    complement(2392. .2410)
FT                   /*tag= aj
FT   primer_bind     complement(2470. .2489)
FT                   /*tag= al
FT   primer_bind     complement(2483. .2500)
FT                   /*tag= am
FT   primer_bind     complement(2563. .2580)
FT                   /*tag= ak
FT   primer_bind     3455. .3474
FT                   /*tag= an
FT   exon            3540. .3884
FT                   /*tag= ao
FT                   /number= 2
FT   misc_binding    3759. .3777
FT                   /*tag= ap
FT   primer_bind     3775. .3792
FT                   /*tag= aq
FT   allele          replace(3778,T)
FT                   /*tag= ar
FT   misc_binding    complement(3779. .3797)
FT                   /*tag= as
FT   primer_bind     complement(3882. .3901)
FT                   /*tag= at
FT   primer_bind     complement(4336. .4356)
FT                   /*tag= au
FT   primer_bind     4444. .4463
FT                   /*tag= av
FT   allele          replace(4498,G)
FT                   /*tag= aw
FT   primer_bind     complement(4902. .4920)
FT                   /*tag= ax
FT   misc_binding    4979. .4997
FT                   /*tag= ay
FT   misc_binding    complement(4999. .5017)
FT                   /*tag= az
FT   misc_binding    5122. .5140
FT                   /*tag= ba
FT   misc_binding    complement(5142. .5160)
FT                   /*tag= bb
FT   primer_bind     6638. .6655
FT                   /*tag= bc
FT   primer_bind     complement(7072. .7089)
FT                   /*tag= bd
FT   misc_binding    7409. .7427
FT                   /*tag= be
FT   misc_binding    complement(7429. .7447)
FT                   /*tag= bf
FT   primer_bind     7995. .8012
FT                   /*tag= bg
FT   misc_binding    8375. .7393
FT                   /*tag= bh
FT   allele          replace(8394,C)
FT                   /*tag= bi
FT   misc_binding    complement(8395. .9413)
FT                   /*tag= bj
FT   primer_bind     complement(8576. .8593)
FT                   /*tag= bk
FT   misc_binding    8685. .8703
FT                   /*tag= bl
FT   allele          replace(8704,T)
FT                   /*tag= bm
FT   misc_binding    complement(8705. .8723)
FT                   /*tag= bn
FT   misc_binding    9009. .9027
FT                   /*tag= bo
FT   allele          replace(9028,G)
FT                   /*tag= bp
FT   misc_binding    complement(9029. .9047)
FT                   /*tag= bq
FT   primer_bind     9622. .9639
FT                   /*tag= br
FT   misc_binding    9931. .9949
FT                   /*tag= bs
FT   misc_binding    complement(9951. .9969)
FT                   /*tag= bt
FT   misc_binding    9958. .9976
FT                   /*tag= bu
FT   primer_bind     9964. .9981
FT                   /*tag= bv
FT   allele          replace(9977,T)
FT                   /*tag= bw
FT   misc_binding    complement(9978. .9996)
FT                   /*tag= bx
FT   misc_binding    10002. .10020
FT                   /*tag= by
FT   allele          replace(10021,G)
FT                   /*tag= bz
FT   misc_binding    complement(10022. .10040)
FT                   /*tag= ca
FT   primer_bind     complement(10023. .10040)
FT                   /*tag= cb
FT   primer_bind     10492. .10512
FT                   /*tag= cc
FT   primer_bind     complement(10546. .10563)
FT                   /*tag= cd
FT   primer_bind     complement(10996. .11015)
FT                   /*tag= cf
FT   misc_binding    11857. .11877
FT                   /*tag= ce
FT   allele          replace(11878,T)
FT                   /*tag= cg
FT   misc_binding    complement(11879. .11897)
FT                   /*tag= ch
FT   primer_bind     11972. .11990
FT                   /*tag= ci
FT   primer_bind     12005. .12023
FT                   /*tag= cj
FT   exon            12163. .12282
FT                   /*tag= ck
FT                   /number= 3
FT   primer_bind     complement(12417. .12436)
FT                   /*tag= cl
```

```
  Query Match             34.8%;  Score 313.6;  DB 4;  Length 23187;
  Best Local Similarity   62.4%;  Pred. No. 1.2e-58;
  Matches  545;  Conservative    0;  Mismatches  319;  Indels   10;  Gaps    3;

Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88
              ||||||| ||| |||| | |||||| | ||| ||||||| || ||||||||||||||||||
Db       5595 TGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA 5654

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
              | |||||| || ||||| |||||||||||| ||| ||||||||||||  ||||||||||
Db       5655 CCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAA 5714

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
              |||||||| ||||||  ||  | |||   |||||||||| |||||||||||| |||||||
Db       5715 AATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 5774

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
              |||||| |||||| |||||||||  ||||||||||||||||||||||| |||||| |||||
Db       5775 GAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCACCA 5834

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
              ||| ||| ||| |||||||||| || |||||||| |||||||||| ||| ||| ||||
Db       5835 CTGTACTGCAGTCTGGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 5894

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
              |     ||  | |  ||  | | |   | |     | |    ||       |||
Db       5895 GACTCCGTCAAGGTATAAGAATGTCAGAGAGTACTAAGTGTTGCAAAG--------AAAA 5946

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
              ||     |     | |  | ||  |  | ||||||  ||       | |    |   |
Db       5947 TAACACCAGGCTGGGTGCATTGGCTCATGCCTGTAAATTTCAGCACTTTGGGAGGCCAAG 6006

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
```

```
Db       6007 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACAACAAAATGAGACCCC 6066

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568

Db       6067 ATGTCTACAAAAATTTTAAAAATTTAAAAATTAGCTGGGCATGGTGGCATGTGCCTGTGG 6126

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628

Db       6127 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 6186

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688

Db       6187 AGTGAGTCATGATCGTGCCACTGCATTCCAGCCTGGGTGA-CAGAGTGAGACCCTGTCTT 6245

Qy        689 TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGG 747

Db       6246 GAAATGAAAAGAAAATAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 6305

Qy        748 AGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGAT 807

Db       6306 AGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCAGCCTGGCCAACATGGT 6365

Qy        808 GAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867

Db       6366 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGGCGTGGTGGTGGGCGCCTGTAA 6425

Qy        868 TCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901

Db       6426 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 6459
```

RESULT 8
ABL67924
ID ABL67924 standard; DNA; 41936 BP.
XX
AC ABL67924;
XX
DT 15-MAY-2002 (first entry)
XX
DE Ovary cancer related gene sequence SEQ ID NO:6261.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 6261; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 41936 BP; 9311 A; 11171 C; 11849 G; 9605 T; 0 U; 0 Other;

Query Match   34.8%; Score 313.6; DB 6; Length 41936;
Best Local Similarity   62.4%; Pred. No. 1.3e-58;
Matches 545; Conservative   0; Mismatches 319; Indels   10; Gaps   3;

```
Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88

Db       1336 TGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA 1395

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148

Db       1396 CCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAA 1455

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208

Db       1456 AATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 1515
```

```
Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
Db       1516 GAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCACCA 1575

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
Db       1576 CTGTACTGCAGTCTGGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAA 1635

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
Db       1636 GACTCCGTCAAGGTATAAGAATGTCAGAGAGTACTAAGTGTTGCAAAG--------AAAA 1687

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
Db       1688 TAACACCAGGCTGGGTGCATTGGCTCATGCCTGTAAATTTCAGCACTTTGGGAGGCCAAG 1747

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
Db       1748 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACAACAAAATGAGACCCC 1807

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
Db       1808 ATGTCTACAAAAATTTTAAAAATTTAAAAATTAGCTGGGCATGGTGGCATGTGCCTGTGG 1867

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
Db       1868 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 1927

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
Db       1928 AGTGAGTCATGATCGTGCCACTGCATTCCAGCCTGGGTGA-CAGAGTGAGACCCTGTCTT 1986

Qy        689 TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGG 747
Db       1987 GAAATGAAAAGAAAATAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 2046

Qy        748 AGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGAT 807
Db       2047 AGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCAGCCTGGCCAACATGGT 2106

Qy        808 GAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867
Db       2107 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGGCGTGGTGGTGGGCGCCTGTAA 2166

Qy        868 TCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
Db       2167 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 2200
```

RESULT 9
ADT25932
ID ADT25932 standard; DNA; 56580 BP.
XX
AC ADT25932;
XX
DT 30-DEC-2004 (first entry)
XX
DE Wild type human ATP-gated ion channel P2X7R genomic DNA.
XX
KW antidepressant; tranquilizer; antimanic; neuroleptic;
KW ATP-gated ion channel P2XR7 modulator; P2X7R; ATP-gated ion channel;
KW affective disorder; depression; dysthymia; atypical depression;
KW premenstrual dysphoric disorder; seasonal affective disorder;
KW anxiety disorder; panic disorder; phobia; agoraphobia; social phobia;
KW obsessive-compulsive disorder; post-traumatic stress disorder;
KW separation anxiety disorder; mania; hypomania; cyclothymic disorder;
KW bipolar disorder; human; gene; ds; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT variation   complement(362)
FT   /*tag= a
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(532)
FT   /*tag= b
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1100)
FT   /*tag= c
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1122)
FT   /*tag= d
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1171)
FT   /*tag= e
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1351)
FT   /*tag= f
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1702)
FT   /*tag= g
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1731)
FT   /*tag= h
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(1860)
FT   /*tag= i
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(2162)
FT   /*tag= j
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(2238)
FT   /*tag= k
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(2373)
FT   /*tag= l
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(2569)
FT   /*tag= m
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(2702)
FT   /*tag= n
FT   /standard_name= "Single nucleotide polymorphism"
FT CDS   3000. .54889
FT   /*tag= p
FT   /product= "Residues 1-328 of P2X7R"
FT   /note= "ATP-gated ion channel"
FT   /transl_except= (pos:47229. .27240, aa:GGKF)
FT exon   3000. .3124
FT   /*tag= o
FT   /number= 1
FT intron   3125. .24840
FT   /*tag= q
FT   /number= 1
FT variation   complement(3166)
FT   /*tag= r
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(24778)
FT   /*tag= s
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   complement(24830)
FT   /*tag= t
FT   /standard_name= "Single nucleotide polymorphism"
FT exon   24841. .25009
FT   /*tag= u
FT   /number= 2
FT variation   complement(24942)
FT   /*tag= v
FT   /standard_name= "Single nucleotide polymorphism"
FT intron   25010. .26133
FT   /*tag= w
FT   /number= 2
FT exon   26134. .26202
FT   /*tag= x
FT   /number= 3

```
FT   variation       complement(26188)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          26203. .30957
FT                   /*tag=  z
FT                   /number= 3
FT   variation       complement(26308)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(26422)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            30958. .31030
FT                   /*tag=  ac
FT                   /number= 4
FT   intron          31031. .32480
FT                   /*tag=  ad
FT                   /number= 4
FT   variation       complement(32394)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(32434)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            32481. .32577
FT                   /*tag=  ag
FT                   /number= 5
FT   variation       complement(32493)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(32506)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(32507)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(32548)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          32578. .35415
FT                   /*tag=  al
FT                   /number= 5
FT   variation       complement(32783)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35309)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35374)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35378)
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            35416. .35496
FT                   /*tag=  aq
FT                   /number= 6
FT   variation       complement(35438)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35454)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          35497. .36112
FT                   /*tag=  at
FT                   /number= 6
FT   variation       complement(35549)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35641)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(35725)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36001)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36064)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36108)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            36113. .36242
FT                   /*tag=  ba
FT                   /number= 7
FT   intron          36243. .37540
FT                   /*tag=  bb
FT                   /number= 7
FT   variation       complement(36374)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36378)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36387)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(36398)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(37439)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       complement(37513)
```

Query Match 34.8%; Score 313.6; DB 13; Length 56580;
Best Local Similarity 62.3%; Pred. No. 1.4e-58;
Matches 574; Conservative 0; Mismatches 319; Indels 29; Gaps 4;

```
Qy          9 TTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGA 68
              |||| |||       |   | ||||||||||| |||| |||| ||| | ||  |||||||
Db      48576 TTAAATAATTGGCCCAAGGCTGGGCACAGTGGCTCACGCCTATAATCCCAACACTTTGGG 48635

Qy         69 AGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGT 128
              || || |||   ||  |||||||||||| |  ||||||||||||||||||| ||| |||||
Db      48636 AGGCCGAGGCAGGCAGATCACTTGAGGTCAAGAGTTCGAGACCAGCCTGGCCAACATGGT 48695

Qy        129 GAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAG 188
              ||||||| |||||||||||||||||||||||||||  || || ||| |||||||||| ||
Db      48696 GAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGAAG 48755

Qy        189 TCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 248
              |||||||||||| |||||||||||||| |||||||||||||||||||| ||||||||||||
Db      48756 TCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGTAGGCAGAGGTTGC 48815

Qy        249 AGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCA 308
              ||||||| ||||||    ||| |||||||||||||| | |||| || || ||||| |||||
Db      48816 AGTGAGCTGAGATCATGCCATTGCACTCCAGCCTGAGTGACAGAGGGAAACTCTGTCTCA 48875

Qy        309 AAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAG 368
              |||||| ||| ||| |||| |  |          || |  |    |  | | |  |
Db      48876 AAAAAAAAAAAAAAAAAAAAATTATAGGACTTGACAAGAGCAGATCTCCAAGACATTTTG 48935

Qy        369 GTAA----------------GCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGC 411
               |||                | |||   |  ||  |||||||  ||| || |||||
Db      48936 TTAAAAGAAAAAAAGCAAACTGTAGAACAATATATGTTATATGATATATAGGTTAAAAAA 48995

Qy        412 ATTTTA----------ACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCC 461
              |  | |          |  ||||||  | |||  |    ||||      |   | |
Db      48996 AAATCACTAGACACAGAAGCTAATCTGTATATTTTCTGCATGTATATTCTATATATACAC 49055

Qy        462 AAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTA 521
              ||  |||     ||     | |  | ||| |||| | |    | | |           |
Db      49056 AAGTGTACACACACACACATACATATACATATGTGTGGGTGTATATATGTACACAAAA 49115
```

```
Qy        522 TAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACT 581
              |  | | | |  |      |    ||   | ||         || |     |      | |
Db      49116 TTGTACCAGTGGCTGCTTCTGGAAAGGAGCTTAGGGTGGGGGAGTAGGAGTAGTCAAAGA 49175

Qy        582 AGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTA 641
              |||         |  | |   ||      |  |    || |  | ||| || |    |
Db      49176 GATTTTAGCGTCATCTGTATTGTTTTGATTTGATTAATTCAGACTTTATCAAGCAGGTCC 49235

Qy        642 TGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAA 701
              | |    |   | | | | || || |  |        | | | |||  | |||
Db      49236 TCTGCGTTCAACTCCATGATTTTCCCCAAAGATAAATCTCTGGTACCTAAAAACAAAGAC 49295

Qy        702 ACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGGCCGAGGCAAA 760
              | || | | | ||||||||||||||||||| |||||| |||||||||||||  |||
Db      49296 GATTGGCTAGACGTGGTGGCTCACGCCTGTAGTCCCAGTACTTTGGGAGGCCAAGGAGGG 49355

Qy        761 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTC 820
              |||||||| ||||||| ||||||  | |||||||||| |||||||  ||||| || ||||
Db      49356 TGGATCACCTGAGGTCAGGAGTTCAAAACCAGCCTGGCCAACATGGAGAAACCCCGTCTC 49415

Qy        821 TAATAAAAATAC-AAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTC 879
              || ||||||||| ||||||||||  |||||||||||     |||||| |||||||| |||
Db      49416 TACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCGCATGCCTGTAATCCCAGCTTCTC 49475

Qy        880 AGGAGTCTGAGGCAGGAGTATC 901
                |||  |||||||||||| |||
Db      49476 TAGAGGCTGAGGCAGGAGAATC 49497
```

```
RESULT 10
ADU23759
ID   ADU23759 standard; DNA; 56580 BP.
XX
AC   ADU23759;
XX
DT   13-JAN-2005  (first entry)
XX
DE   Human ATP-gated ion channel protein (P2X7R) gene - SEQ ID 1.
XX
KW   ATP-gated ion channel protein; P2X7R; affective disorder;
KW   major depression; dysthymia; atypical depression;
KW   premenstrual dysphoric disorder; seasonal affective disorder;
KW   generalised anxiety disorder; panic disorder; phobia; agoraphobia;
KW   social phobia; specific phobia; obsessive-compulsive disorder;
KW   post-traumatic stress disorder; separation anxiety disorder; mania;
KW   hypomania; cyclothymic disorder; type I bipolar disorder;
KW   type II bipolar disorder; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3000. .54889
FT                   /*tag=  b
FT                   /product= "Human ATP-gated ion channel protein (P2X7R) -
FT                   SEQ ID 10"
FT   exon            3000. .3124
FT                   /*tag=  a
FT                   /number= 1
FT   intron          3125. .24840
FT                   /*tag=  c
FT                   /number= 1
FT   exon            24841. .25009
FT                   /*tag=  d
FT                   /number= 2
FT   intron          25010. .26133
FT                   /*tag=  e
FT                   /number= 2
FT   exon            26134. .26202
FT                   /*tag=  f
FT                   /number= 3
FT   intron          26203. .30957
FT                   /*tag=  g
FT                   /number= 3
FT   exon            30958. .31030
FT                   /*tag=  h
FT                   /number= 4
FT   intron          31031. .32480
FT                   /*tag=  i
FT                   /number= 4
FT   exon            32481. .32577
FT                   /*tag=  j
FT                   /number= 5
FT   intron          32578. .35415
FT                   /*tag=  k
FT                   /number= 5
FT   exon            35416. .35496
FT                   /*tag=  l
FT                   /number= 6
FT   intron          35497. .36112
FT                   /*tag=  m
FT                   /number= 6
FT   exon            36113. .36242
FT                   /*tag=  n
FT                   /number= 7
FT   intron          36243. .37540
FT                   /*tag=  o
FT                   /number= 7
FT   exon            37541. .37677
FT                   /*tag=  p
FT                   /number= 8
FT   intron          37678. .45469
FT                   /*tag=  q
FT                   /number= 8
FT   exon            45470. .45560
FT                   /*tag=  r
FT                   /number= 9
FT   intron          45561. .47229
FT                   /*tag=  s
FT                   /number= 9
FT   exon            47230. .47295
FT                   /*tag=  t
FT                   /number= 10
FT   intron          47296. .47379
FT                   /*tag=  u
FT                   /number= 10
FT   exon            47380. .47529
FT                   /*tag=  v
FT                   /number= 11
FT   intron          47530. .50437
FT                   /*tag=  w
FT                   /number= 11
FT   exon            50438. .50539
FT                   /*tag=  x
FT                   /number= 12
FT   intron          50540. .54391
FT                   /*tag=  y
FT                   /number= 12
FT   exon            54392. .54889
FT                   /*tag=  z
FT                   /number= 13
XX
PN   EP1473367-A1.
XX
PD   03-NOV-2004.
XX
PF   16-APR-2004; 2004EP-00018640.
XX
PR   17-APR-2003; 2003EP-00008753.
PR   04-SEP-2003; 2003EP-00019626.
PR   16-APR-2004; 2004EP-00009160.
XX
PA   (NEUR-) NEURONOVA AG.
XX
PI   Barden N,  Sillaber I,  Paez-Pereda M;
XX
```

DR WPI; 2004-787080/78.
DR P-PSDB; ADU23768.
XX
PT Use of nucleic acid molecules comprising sequence encoding P2X7R with a
PT mutation in 5'UTR, 3'UTR, exon 3, 5, 6, 8 or 13, introns 1, 3, 4 - 7, 9,
PT 11 or 12 or deletion of exon 13 for preparation of composition for
PT detecting affective disorder.
XX
PS Claim 1; SEQ ID NO 1; 188pp; English.
XX
CC The invention comprises the amino acid and coding sequences of a human
CC ATP-gated ion channel protein (P2X7R), the invention also comprises
CC allele-specific PCR primers for the human P2X7R gene. The DNA and protein
CC sequences of the invention are useful in the preparation of an affective
CC disorder, such as: major depression (e.g. dysthymia, atypical depression,
CC premenstrual dysphoric disorder and seasonal affective disorder),
CC generalised anxiety disorder (e.g. panic disorder, phobias, agoraphobia,
CC social phobia, specific phobia, obsessive-compulsive disorder, post-
CC traumatic stress disorder, separation anxiety disorder, mania, hypomania
CC and cyclothymic disorder), and bipolar disorder (e.g. type I or type II
CC bipolar disorder). The present DNA sequence represents the human P2X7R
CC gene of the invention.
XX
SQ Sequence 56580 BP; 15379 A; 13163 C; 13083 G; 14955 T; 0 U; 0 Other;

Query Match 34.8%; Score 313.6; DB 13; Length 56580;
Best Local Similarity 62.3%; Pred. No. 1.4e-58;
Matches 574; Conservative 0; Mismatches 319; Indels 29; Gaps 4;

```
Qy        9 TTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGA 68
Db    48576 TTAAATAATTGGCCCAAGGCTGGGCACAGTGGCTCACGCCTATAATCCCAACACTTTGGG 48635

Qy       69 AGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGT 128
Db    48636 AGGCCGAGGCAGGCAGATCACTTGAGGTCAAGAGTTCGAGACCAGCCTGGCCAACATGGT 48695

Qy      129 GAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAG 188
Db    48696 GAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGAAG 48755

Qy      189 TCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 248
Db    48756 TCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGTAGGCAGAGGTTGC 48815

Qy      249 AGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCA 308
Db    48816 AGTGAGCTGAGATCATGCCATTGCACTCCAGCCTGAGTGACAGAGGGAAACTCTGTCTCA 48875

Qy      309 AAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAG 368
Db    48876 AAAAAAAAAAAAAAAAAAAAAATTATAGGACTTGACAAGAGCAGATCTCCAAGACATTTTG 48935

Qy      369 GTAA----------GCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGC 411
Db    48936 TTAAAAGAAAAAAAGCAAACTGTAGAACAATATATGTTATATGATATATAGGTTAAAAAA 48995

Qy      412 ATTTTA----------ACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCC 461
Db    48996 AAATCACTAGACACAGAAGCTAATCTGTATATTTTCTGCATGTATATTCTATATATACAC 49055

Qy      462 AAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTA 521
Db    49056 AAGTGTACACACACACACACATACATATACATATGTGTGGGTGTATATATGTACACAAAA 49115

Qy      522 TAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACT 581
Db    49116 TTGTACCAGTGGCTGCTTCTGGAAAGGAGCTTAGGGTGGGGGAGTAGGAGTAGTCAAAGA 49175

Qy      582 AGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTA 641
Db    49176 GATTTTAGCGTCATCTGTATTGTTTTGATTTGATTAATTCAGACTTTATCAAGCAGGTCC 49235

Qy      642 TGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAA 701
Db    49236 TCTGCGTTCAACTCCATGATTTTCCCCAAAGATAAATCTCTGGTACCTAAAAACAAAGAC 49295

Qy      702 ACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGGCCGAGGCAAA 760
Db    49296 GATTGGCTAGACGTGGTGGCTCACGCCTGTAGTCCCAGTACTTTGGGAGGCCAAGGAGGG 49355

Qy      761 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTC 820
Db    49356 TGGATCACCTGAGGTCAGGAGTTCAAAACCAGCCTGGCCAACATGGAGAAACCCCGTCTC 49415

Qy      821 TAATAAAAATAC-AAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTC 879
Db    49416 TACTAAAAATACAAAAAATTAGCCAGGCATGGTGGCGCATGCCTGTAATCCCAGCTTCTC 49475

Qy      880 AGGAGTCTGAGGCAGGAGTATC 901
Db    49476 TAGAGGCTGAGGCAGGAGAATC 49497
```

RESULT 11
AAX83427
ID AAX83427 standard; DNA; 21721 BP.
XX
AC AAX83427;
XX
DT 31-AUG-1999 (first entry)
XX
DE Human lipolysis stimulated receptor genomic sequence.
XX
KW Human; lipolysis stimulated receptor; LSR; lipoprotein; free fatty acid;
KW cytokine; probe; primer; amplification; hybridisation; detection; leptin;
KW allele; variant; mutation; deletion; loss of heterozygosity; chylomicron;
KW transgenic animal; gene expression; triglyceride; eating disorder;
KW obesity; atheromatosis; atherosclerosis; hypertension; diabetes;
KW anorexia; metabolism; ds.
XX
OS Homo sapiens.
XX
PN WO9907737-A2.
XX
PD 18-FEB-1999.
XX
PF 06-AUG-1998; 98WO-IB001257.
XX
PR 06-AUG-1997; 97FR-00010088.
PR 22-APR-1998; 98FR-00005032.
XX
PA (GEST) GENSET.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bihain B, Bougueleret L, Yen-Potin F;
XX
DR WPI; 1999-190035/16.
XX
PT Lipolysis stimulated receptor involved in leptin metabolism - and
PT controlling its activity for treatment of obesity, anorexia or abnormal
PT cytokine metabolism.
XX
PS Claim 40; Page 258-277; 279pp; French.
XX
CC This sequence represents the human lipolysis stimulated receptor (LSR)
CC genomic sequence. The novel LSR binds lipoproteins in presence of free
CC fatty acids (FFA) and binds cytokines in absence of FFA. The nucleic acid
CC and its fragments are used as probes and primers for detection and/or
CC amplification of LSR genes; for production of recombinant LSR; for
CC detecting allelic variants, mutations, deletions, loss of heterozygosity
CC and genetic abnormalities in the gene. LSR, recombinant cells and
CC transgenic animals are used to screen for chemical interacting with LSR,
CC also to study expression and activity of LSR and its interactions. The
CC chemicals, and leptin, are used to modulate the number of LSR in a cell,
CC its recycling rate and/or specificity of receptor activity, particularly

CC for reducing the level of leptin, lipoproteins, chylomicrons and/or
CC triglycerides. The chemicals are thus useful for treating eating
CC disorders, particularly obesity (and related diseases such as
CC atheromatosis, atherosclerosis, hypertension and diabetes) or anorexia,
CC also disease associated with abnormal cytokine metabolism
XX
SQ Sequence 21721 BP; 4980 A; 5688 C; 6170 G; 4865 T; 0 U; 18 Other;

Query Match 34.7%; Score 312.6; DB 2; Length 21721;
Best Local Similarity 62.2%; Pred. No. 1.9e-58;
Matches 544; Conservative 0; Mismatches 320; Indels 10; Gaps 3;

```
Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88
              ||||||| ||| |||| | |||||| | ||| ||||||| || ||||||||||||||||
Db       5494 TGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGGATCA 5553

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
              | |||||| || ||||| |||||||||||| ||| |||||||||||  ||||||||||
Db       5554 CCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAA 5613

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
              |||||||| |||||| ||  | |||  |||||||||| |||||||||||  |||||||
Db       5614 AATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCT 5673

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
              |||||| |||||| |||||||||  ||||||||||||||||||||||| |||||| |||||
Db       5674 GAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATCGCACCA 5733

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
              ||| ||| |||  ||||||||| || |||||||| |||||||||| ||| ||| ||||
Db       5734 CTGTACTGCAGTNTGGGCGACAGAGTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 5793

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
              |         ||  | |  ||  | | |   | |    ||       |||
Db       5794 GACTCCGTCAAGGTATAAGAATGTCAGAGAGTACTAAGTGTTGCAAAG--------AAAA 5845

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
              ||     |   | |  | ||  |  |  ||||||  ||      | |    |    |
Db       5846 TAACACCAGGCTGGGTGCATTGGCTCATGCCTGTAAATTTCAGCACTTTGGGAGGCCAAG 5905

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
                | |   |  ||   | |  || | ||    |   | |   | | |
Db       5906 GCAGGAGGATCACTTGAGCCTAGGAGTTTGAGACCAGCCTGGACAACAAAATGAGACCCC 5965

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
                |   ||  ||  |   | |  |||       |       |                |
Db       5966 ATGTCTACAAAAATTTTAAAAATTTAAAAATTAGCTGGGCATGGTGGCATGTGCCTGTGG 6025

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
              |   | | | |   |     |  | | |     |   | ||  |        ||
Db       6026 TCCCGGCTGCTCAGGAGGCTGAGGTGGGAGGATTGCTTGGGCTTGAGAGGTCAAGGCTTC 6085

Qy        629 TTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCT 688
                |    |   ||  | | ||||  ||  ||    |  |   | | |     | |
Db       6086 AGTGAGTCATGATCGTGCCACTGCATTCCAGCCTGGGTGA-CAGAGTGAGACCCTGTCTT 6144

Qy        689 TTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGG 747
                || | | ||||| |||| |||||   ||||||||| ||||| ||||||||||||||||
Db       6145 GAAATGAAAAGAAAATAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 6204

Qy        748 AGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGAT 807
              |||||||||    |||||||| ||||||| |||| | ||||||||||||| |||||||| |
Db       6205 AGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGATCGAGACCAGCCTGGCCAACATGGT 6264

Qy        808 GAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAG 867
              ||||  || |||||| |||||||||||||| ||||| |||| |||||||  || ||||||
Db       6265 GAAATCCCATCTCTACTAAAAATACAAAATTTAGCCGGGCGTGGTGGTGGGCGCCTGTAA 6324

Qy        868 TCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              |||||||||||| |||| |||||||||||| |||
Db       6325 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATC 6358
```

RESULT 12
ADS19371/c
ID ADS19371 standard; DNA; 37940 BP.
XX
AC ADS19371;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human PIP5K type II beta related DNA #1.
XX
KW Phosphatidylinositol-4-phosphate 5-kinase type II-beta; PIP5KII-beta;
KW hyperproliferative disorder; cancer; inflammation; metabolic disorder;
KW angiogenesis; therapy; human; ds.
XX
OS Homo sapiens.
XX
PN US2003232777-A1.
XX
PD 18-DEC-2003.
XX
PF 16-JAN-2003; 2003US-00348073.
XX
PR 18-JUN-2002; 2002US-00175627.
XX
PA (MARC/) MARCUSSON E G.
PA (DOBI/) DOBIE K W.
PA (FREI/) FREIER S M.
XX
PI Marcusson EG, Dobie KW, Freier SM;
XX
DR WPI; 2004-070378/07.
DR GENBANK; AC006449.19.
XX
PT New compound targeted to a nucleic acid molecule encoding
PT phosphatidylinositol-4-phosphate 5-kinase, type II beta, useful in
PT treating hyperproliferative disorder, e.g. cancer, inflammatory or
PT metabolic disorder.
XX
PS Example 13; SEQ ID NO 12; 67pp; English.
XX
CC The present invention is directed to antisense oligonucleotides which are
CC targetted to a nucleic acid encoding phosphatidylinositol-4-phosphate 5-
CC kinase type II (PIP5KII) beta and which modulate the expression of
CC PIP5KII. The invention is useful for treating a disease or condition
CC associated with phosphatidylinositol-4-phosphate 5-kinase type II beta
CC such as hyperproliferative disorder (cancer), inflammatory, metabolic
CC disorder and inhibiting angiogenesis. The present sequence is human
CC PIP5KII related DNA.
XX
SQ Sequence 37940 BP; 8915 A; 9323 C; 9359 G; 10343 T; 0 U; 0 Other;

Query Match 34.7%; Score 312.4; DB 13; Length 37940;
Best Local Similarity 64.3%; Pred. No. 2.4e-58;
Matches 561; Conservative 0; Mismatches 281; Indels 30; Gaps 5;

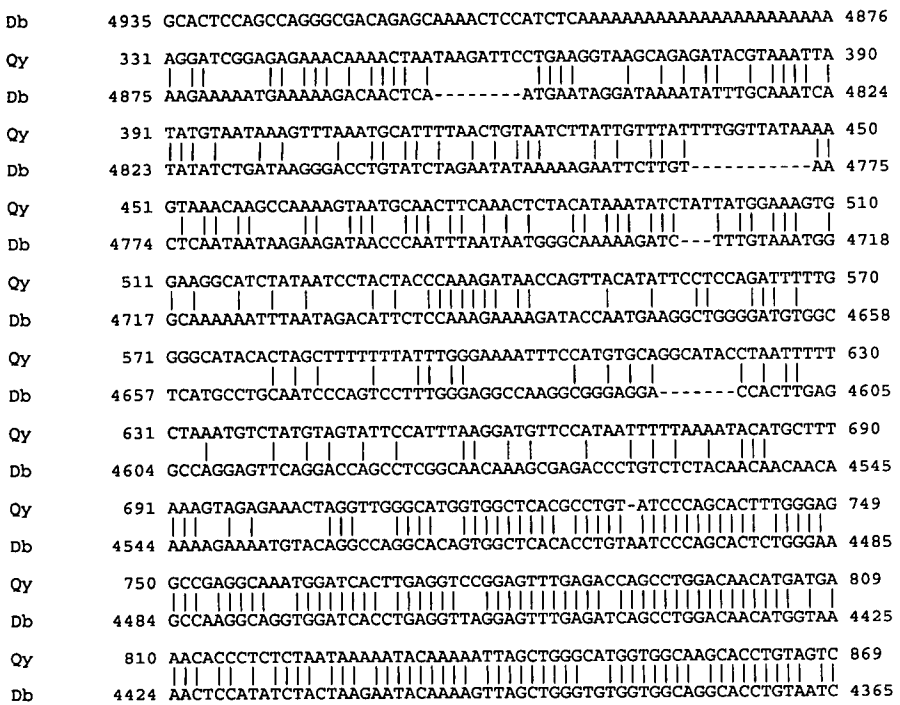
```
Qy         31 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 90
              ||  ||||| |||| |||||||| | ||  ||||||| || || ||| ||| |||| ||
Db       5175 GGTGCAGTGGCTCACGCCTGTAATCCCACCACTTTGGGAGGCCGAGGCGGGGGGATTGCT 5116

Qy         91 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 150
              |||||  || |||||||||||||| |||| ||| ||||  ||||||| ||||||||||||
Db       5115 TGAGGCCAGGAGTTCGAGACCAGTCTGGCCAACATGGCAAAACCCTGTCTCTACTAAAAA 5056

Qy        151 TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGA 210
              |||||||||||||  ||||| |||  |  |||||| |||||| ||||| |||||||||
Db       5055 TACAAAAATTAGCCTGGTGTGGTGGCGGGCACCTGTAATCCCAGTTACTCAGGAGGCTGA 4996

Qy        211 GGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACT 270
              |||| |||||| ||||||||| |||||||||||||||||||||| ||||| ||||  |
Db       4995 GGCAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCGTT 4936

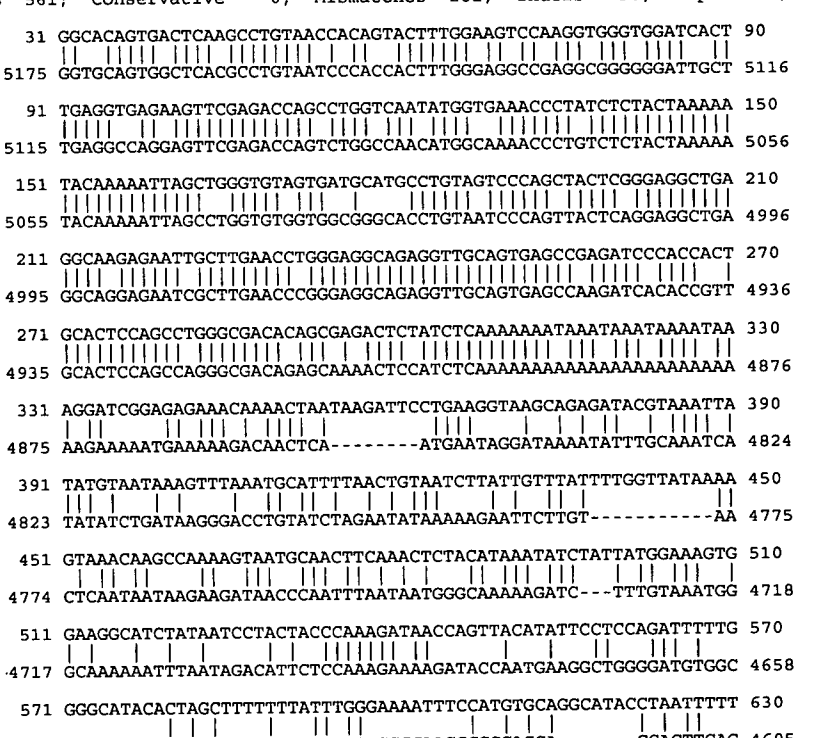
Qy        271 GCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 330
              |||||||||||  |||||||| ||| | |||| |||||||||||| ||| ||| |||| ||
```



RESULT 13
ADJ26093/c
ID ADJ26093 standard; DNA; 38000 BP.
XX
AC ADJ26093;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human PIP5KII-beta genomic DNA SeqID 11.
XX
KW human; PIP5KII-beta;
KW phosphatidylinositol 4-phosphate 5-kinase, type II beta;
KW cell differentiation; proliferation; motility; cancer; inflammatory;
KW metabolic disease; infection; tumour formation; cytostatic;
KW antiinflammatory; ds.
XX
OS Homo sapiens.
XX
PN US2003232775-A1.
XX
PD 18-DEC-2003.
XX
PF 18-JUN-2002; 2002US-00175627.
XX
PR 18-JUN-2002; 2002US-00175627.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Dobie KW, Freier SM;
XX

DR WPI; 2004-061311/06.
XX
PT New antisense compound targeted to a nucleic acid molecule encoding
PT phosphatidylinositol-4-phosphate 5-kinase, type II, useful for modulating
PT expression of phosphatidylinositol-4-phosphate 5-kinase, type II or
PT treating cancer.
XX
PS Example 15; SEQ ID NO 11; 63pp; English.
XX
CC This invention relates to novel compositions and methods for modulating
CC the expression of PIP5KII-beta (also known as phosphatidylinositol 4-
CC phosphate 5-kinase, type II beta). Specifically, it refers to antisense
CC compounds that can target and hybridise with a nucleic acid molecule
CC encoding PIP5KII-beta, as well as splice variants thereof, to inhibit
CC expression accordingly. PIP5KII-beta is a homodimeric enzyme that is a
CC member of the superfamily of phosphatidylinositolphosphate kinases and is
CC thought to play a role in pathological conditions in which cell
CC differentiation, proliferation and/ or motility are compromised, such as
CC cancer, inflammatory disorders and metabolic diseases. The present
CC invention describes antisense compounds that can be used as research
CC tools in differential and/ or combinatorial analyses to elucidate
CC expression patterns of a portion or the entire complement of genes
CC expressed within cells and tissues. The compound can also be used as
CC prophylaxis, e.g. to prevent or delay infection, inflammation or tumour
CC formation. Accordingly, these compounds can be described as having
CC cytostatic and antiinflammatory activities. This polynucleotide sequence
CC is the human PIP5KII-beta genomic DNA of the invention.
XX
SQ Sequence 38000 BP; 8927 A; 9338 C; 9374 G; 10356 T; 0 U; 5 Other;

Query Match 34.7%; Score 312.4; DB 12; Length 38000;
Best Local Similarity 64.3%; Pred. No. 2.4e-58;
Matches 561; Conservative 0; Mismatches 281; Indels 30; Gaps 5;



```
Qy        631 CTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTT 690
              |     | | | |  |     || | |     |  | | | |  |||
Db       4604 GCCAGGAGTTCAGGACCAGCCTCGGCAACAAAGCGAGACCCTGTCTCTACAACAACAACA 4545

Qy        691 AAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGGAG 749
              |||  | |       |||   ||||  ||||||||| ||||| |||||||||| |||||
Db       4544 AAAAGAAAATGTACAGGCCAGGCACAGTGGCTCACACCTGTAATCCCAGCACTCTGGGAA 4485

Qy        750 GCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGA 809
              ||| |||||  |||||||| ||||||  |||||||||| |||||||||||||| | |
Db       4484 GCCAAGGCAGGTGGATCACCTGAGGTTAGGAGTTTGAGATCAGCCTGGACAACATGGTAA 4425

Qy        810 AACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTC 869
              ||| || | |||| ||| ||||||||| |||||||||  |||||||| ||||||||| ||
Db       4424 AACTCCATATCTACTAAGAATACAAAAGTTAGCTGGGTGTGGTGGCAGGCACCTGTAATC 4365

Qy        870 CCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              | || ||||  |||| ||||||||||||| |||
Db       4364 CGAGGTACTTGGGAGGCTGAGGCAGGAGAATC 4333
```

RESULT 14
AAL35901/c
ID AAL35901 standard; DNA; 28818 BP.
XX
AC AAL35901;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2266.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001338.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal
PT cancers and also for testing and detection e.g. diagnosis.
XX
PS Example 2; SEQ ID NO 2266; 781pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 28818 BP; 7658 A; 5980 C; 6613 G; 8567 T; 0 U; 0 Other;

Query Match 34.4%; Score 310; DB 4; Length 28818;
Best Local Similarity 62.8%; Pred. No. 7.5e-58;

Matches 549; Conservative 0; Mismatches 315; Indels 10; Gaps 4;

```
Qy        37 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 96
             ||| |||| |||||||| | ||| ||||||| || | ||||  |||||||||| |||| |
Db     19755 GTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTAAGGCAGGTGGATCACCTGAGAT 19696

Qy        97 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 156
             || ||||   |||||||||||| ||| ||||||| |||| |||||||||||||| ||||||
Db     19695 CAGGAGTTGAGGACCAGCCTGGCCAACATGGTGAGACCCCATCTCTACTAAAACTACAAA 19636

Qy       157 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 216
             |||||| | || || |||  |  |||||||| |||||||||| |||||| || |||| |
Db     19635 AATTAGGTAGGAGTGGTGTCGGGTGCCTGTAATCCCAGCTACTTGGGAGGTTGGGGCAGG 19576

Qy       217 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 276
             |||||||||||| |  |||| |||||||||| ||| ||||| || | ||| |||||||
Db     19575 AGAATTGCTTGAAGCCAGGAGTCAGAGGTTGCAATGAACCGAGGTCGCGCCATTGCACTC 19516

Qy       277 CAGCCTGGGCGACACAGCGAGACTCTATCTC-AAAAAAATAAATAAATAAAATAAAGGAT 335
             || |||||| |||| || |||||||| |||| ||| |||||||     | |||| |  |||
Db     19515 CACCCTGGGTGACAGAGTGAGACTCTGTCTCTAAATAAATAAACGCTTCAAATCATAGAT 19456

Qy       336 CGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTA---AGCAGAGATACGTAAATTATA 392
             |   ||| |||| | | ||  |       | ||   |  | |||||
Db     19455 AAAATTAAAAGAAAAATTAAAATCTGGAAAGGGATATGTATTACAGATAGGACTGAGGAT 19396

Qy       393 TGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGT 452
             | |  ||    | |||    | |   | ||   || |   |   ||  | | |
Db     19395 TTTCTTACTTATAAAAAAGGCTATTTATATACCAATAATAAGGAAGATGTGAACACAGCA 19336

Qy       453 AAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGA 512
             |||  | | ||     |     |  |  | |    ||  |  ||   | |||  |
Db     19335 AAAATATCAAAGGACATAAATAGAAAATTCATGAGAGTGGTAAAAAAAATTCAAACTCAC 19276

Qy       513 AGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGG 572
             |   | |    ||| |    ||   |  |  ||   | |       | |  |   |
Db     19275 TAATAATCAAAGGAGTACAAATTGAAATGAGACCCTTTTCTGTCTAGAAATTTAATAATG 19216

Qy       573 GCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCT 632
             | |||  | ||  ||    || |  | || |  ||      |  | | |      | |
Db     19215 GAATAATCCAGAATTGGGGATATAAGGAAGTAAACACTCATATACTGATCATGGGAATAT 19156

Qy       633 AAATGTCTATGTA-----GTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGC 687
             |||      |  |     |  | || ||  || |  || || | |||||     |
Db     19155 AAAGTGGCAGAAATCTGCAAAGACACTTGAAATATCTATCAAAAGTCTTAAAGAGAAGCA 19096

Qy       688 TTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGG 746
             |||  |   | || ||||| ||||   | |||||||| ||||| ||||||||||||||
Db     19095 TTTCTAAAGAAAAACCTAGGCTGGGTGCGATGGCTCACACCTGTAATCCCAGCACTTTGG 19036

Qy       747 GAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGA 806
             |||||| |||||| |||||| ||||||  | |||||||||||||| ||||  ||||||
Db     19035 GAGGCCAAGGCAAGTGGATCTCTTGAGCCCAGGAGTTTGAGACCAGGCTGGCTAACATGG 18976

Qy       807 TGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTA 866
             ||||   |||||| |||||||||||||||||||||||||| |||||||| ||||||||
Db     18975 CAAAACCGTGTCTCTACTAAAAATACAAAAATTAGCTGGGCCTGGTGGCACACACCTGTA 18916

Qy       867 GTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
             || ||||||||  |||| ||||| |||||| ||
Db     18915 GTTCCAGCTACTTGGGAGGCTGAGCCAGGAGAAT 18882
```

RESULT 15
ABX58889/c
ID ABX58889 standard; cDNA; 28818 BP.
XX
AC ABX58889;
XX
DT 26-FEB-2003 (first entry)
XX
DE cDNA encoding novel human musculoskeletal system antigen #1233.
XX

KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;
KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;
KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
KW post-operative tissue repair; limb regeneration; neuronal growth;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW AIDS-related complex; chondrocyte growth; bone regeneration;
KW periodontal regeneration; tissue transport; bone graft; skin aging;
KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
KW cell growth; organ transplant; cell differentiation; body height; weight;
KW hair colour; eye colour; skin; percentage of adipose tissue;
KW pigmentation; cosmetic surgery; metabolism; biorhythm; caricadic rhythm;
KW depression; tendency for violence; pain; reproductive capability;
KW hormone level; endocrine level; appetite; libido; memory; stress;
KW storage capability; fat content; lipid content; protein content;
KW carbohydrate content; vitamin content; cofactor content;
KW nutritional component.
XX
OS Homo sapiens.
XX
PN US2002147140-A1.
XX
PD 10-OCT-2002.
XX
PF 17-JAN-2001; 2001US-00764877.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-128199/12.
XX
PT Isolated nucleic acid molecules encoding musculoskeletal system
PT associated polypeptides, useful for detecting disorders, e.g. cancer.
XX
PS Disclosure; SEQ ID NO 2266; 321pp; English.
XX
CC The invention describes an isolated nucleic acid molecule comprising a
CC sequence encoding musculoskeletal system associated polypeptides useful
CC for detecting disorders, e.g., cancer or cancer metastases, in animals or
CC humans. The nucleic acid: stimulates re-vascularisation of ischaemic
CC tissues associated with conditions such as thrombosis, arteriosclerosis,
CC and other cardiovascular conditions; treats wounds due to injuries,
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
CC and limb regeneration; stimulates neuronal growth; can treat and prevent
CC neuronal damage occurring in certain disorders or neurodegenerative
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-
CC related complex; stimulates chondrocyte growth, thus they can be used to
CC enhance bone and periodontal regeneration and aid in tissue transports or
CC bone grafts; prevents skin aging due to sunburn by stimulating
CC keratinocyte growth; prevents hair loss, since FGF family members
CC activate hair-forming cells and promotes melanocyte growth; stimulates
CC growth and differentiation of hematopoietic cells and bone marrow cells
CC when used in combination with other cytokines; maintains organs before
CC transplantation or for supporting cell culture of primary tissues;
CC induces tissue of mesodermal origin to differentiate in early embryos;
CC increases or decreases the differentiation or proliferation of embryonic
CC stem cells, besides, haematopoietic lineage; modulates mammalian
CC characteristics, such as, body height, weight, hair colour, eye colour,
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal
CC state or physical state by influencing biorhythms, caricadic rhythms,
CC depression, tendency for violence, tolerance for pain, reproductive
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or
CC stress; increases or decreases storage capabilities, fat content, lipid,
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional
CC components. This sequence encodes a novel human musculoskeletal system
CC antigen. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140
XX
SQ Sequence 28818 BP; 7658 A; 5980 C; 6613 G; 8567 T; 0 U; 0 Other;

Query Match 34.4%; Score 310; DB 8; Length 28818;
Best Local Similarity 62.8%; Pred. No. 7.5e-58;
Matches 549; Conservative 0; Mismatches 315; Indels 10; Gaps 4;

Qy 37 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 96
Db 19755 GTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTAAGGCAGGTGGATCACCTGAGAT 19696

Qy 97 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 156
Db 19695 CAGGAGTTGAGGACCAGCCTGGCCAACATGGTGAGACCCCATCTCTACTAAAACTACAAA 19636

Qy 157 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 216
Db 19635 AATTAGGTAGGAGTGGTGTCGGGTGCCTGTAATCCCAGCTACTTGGGAGGTTGGGGCAGG 19576

Qy 217 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 276
Db 19575 AGAATTGCTTGAAGCCAGGAGTCAGAGGTTGCAATGAACCGAGGTCGCGCCATTGCACTC 19516

```
Qy        277 CAGCCTGGGCGACACAGCGAGACTCTATCTC-AAAAAAATAAATAAATAAAATAAAGGAT 335
              || |||||| |||| || |||||||| |||| ||| |||||||    | ||||  |  |||
Db      19515 CACCCTGGGTGACAGAGTGAGACTCTGTCTCTAAATAAATAAACGCTTCAAATCATAGAT 19456

Qy        336 CGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTA---AGCAGAGATACGTAAATTATA 392
                |   ||| |||| | | ||  |         | ||   |  | ||||| |
Db      19455 AAAATTAAAAGAAAAATTAAAATCTGGAAAGGGATATGTATTACAGATAGGACTGAGGAT 19396

Qy        393 TGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGT 452
              | |  ||    | |||      | |   | ||    || |    |   ||   | | |
Db      19395 TTTCTTACTTATAAAAAAGGCTATTTATATACCAATAATAAGGAAGATGTGAACACAGCA 19336

Qy        453 AAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGA 512
              |||  | | ||    |     |    |  |  | |    ||   |  ||  ||| |
Db      19335 AAAATATCAAAGGACATAAATAGAAAATTCATGAGAGTGGTAAAAAAAATTCAAACTCAC 19276

Qy        513 AGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGG 572
                |  | |    ||| |    ||   |  | ||   | |       |  |  |   |
Db      19275 TAATAATCAAAGGAGTACAAATTGAAATGAGACCCTTTTCTGTCTAGAAATTTAATAATG 19216

Qy        573 GCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCT 632
              | |||  | ||  ||    || |  | || |   ||      |  |  | |       | |
Db      19215 GAATAATCCAGAATTGGGGATATAAGGAAGTAAACACTCATATACTGATCATGGGAATAT 19156

Qy        633 AAATGTCTATGTA-----GTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGC 687
              |||     |   |       |  |  || ||  || |  || || | |||||    |
Db      19155 AAAGTGGCAGAAATCTGCAAAGACACTTGAAATATCTATCAAAAGTCTTAAAGAGAAGCA 19096

Qy        688 TTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGG 746
              |||  |    | || ||||| ||||   | ||||||| ||||| ||||||||||||||
Db      19095 TTTCTAAAGAAAAACCTAGGCTGGGTGCGATGGCTCACACCTGTAATCCCAGCACTTTGG 19036

Qy        747 GAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGA 806
              |||||| |||||| |||||| ||||||  | ||||||||||||||| ||||  ||||||
Db      19035 GAGGCCAAGGCAAGTGGATCTCTTGAGCCCAGGAGTTTGAGACCAGGCTGGCTAACATGG 18976

Qy        807 TGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTA 866
                ||||    |||||| |||||||||||||||||||||||| ||||||||  ||||||||
Db      18975 CAAAACCGTGTCTCTACTAAAAATACAAAAATTAGCTGGGCCTGGTGGCACACACCTGTA 18916

Qy        867 GTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              || |||||||||  |||| ||||| ||||||  ||
Db      18915 GTTCCAGCTACTTGGGAGGCTGAGCCAGGAGAAT 18882
```

Search completed: February 9, 2006, 18:54:21
Job time : 584.391 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 15:02:28 ; Search time 3367.06 Seconds
(without alignments)
12519.885 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 901
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	310.4	34.5	507	5	BU785040	BU785040 in43g09.x
c	2	281	31.2	2213	4	CR623485	CR623485 full-leng
c	3	280	31.1	1218	4	CR626853	CR626853 full-leng
c	4	276	30.6	6056	4	CR627178	CR627178 Homo sapi
c	5	260.4	28.9	903	5	BX431110	BX431110 BX431110
	6	253.2	28.1	2495	4	HSM808204	BX648058 Homo sapi
c	7	253	28.1	2097	4	CR614786	CR614786 full-leng
	8	252.4	28.0	1945	4	CR607080	CR607080 full-leng
c	9	250.2	27.8	1809	4	CR604283	CR604283 full-leng
c	10	247	27.4	5468	9	AQ839834	AQ839834 260L13-C5
c	11	244.8	27.2	3737	4	CR749536	CR749536 Homo sapi
	12	244.6	27.1	669	7	CN414771	CN414771 170006008
c	13	243.6	27.0	691	6	CD102556	CD102556 AGENCOURT
	14	243.2	27.0	710	10	AG186254	AG186254 Pan trogl
c	15	242	26.9	4640	4	AF370428	AF370428 Homo sapi
	16	241.4	26.8	4160	4	CR859494	CR859494 Pongo pyg
c	17	240.6	26.7	655	6	CA439825	CA439825 UI-H-DI0-
c	18	239.2	26.5	2288	4	CR857366	CR857366 Pongo pyg
c	19	238.6	26.5	4663	4	CR857085	CR857085 Pongo pyg
	20	238	26.4	608	10	AG161060	AG161060 Pan trogl
c	21	237.8	26.4	695	2	BG566642	BG566642 602585490
c	22	237.6	26.4	788	8	CX782424	CX782424 HESC3_14_
c	23	237.4	26.3	649	11	CR960737	CR960737 Homo sapi
c	24	237	26.3	464	5	BX493340	BX493340 DKFZp781I
	25	236.6	26.3	346	9	AQ706012	AQ706012 HS_5556_A
c	26	236.6	26.3	1114	3	BM467056	BM467056 AGENCOURT
	27	236.2	26.2	369	5	BU536852	BU536852 AGENCOURT
	28	236.2	26.2	539	9	AQ379787	AQ379787 RPCI11-15
	29	235.8	26.2	460	9	B15692	B15692 345N8.TP CI
c	30	235.8	26.2	702	6	CA428680	CA428680 UI-H-DF0-
	31	235.8	26.2	779	9	AQ586732	AQ586732 RPCI-11-4
	32	235.8	26.2	3355	4	CR857235	CR857235 Pongo pyg
	33	235.6	26.1	681	9	AQ543621	AQ543621 RPCI-11-3
	34	235.4	26.1	666	10	AG089378	AG089378 Pan trogl
c	35	235.4	26.1	753	5	BU853329	BU853329 AGENCOURT
	36	235.4	26.1	4146	4	HSM805210	AL833904 Homo sapi
c	37	235.2	26.1	641	5	BX102838	BX102838 BX102838
c	38	235	26.1	563	7	CK902534	CK902534 im66g10.x
c	39	235	26.1	576	5	BU078278	BU078278 im66g10.y
	40	235	26.1	744	8	CX752348	CX752348 AGENCOURT
	41	234.6	26.0	5022	4	CR933599	CR933599 Homo sapi
c	42	234.4	26.0	687	6	CA448731	CA448731 UI-H-EI0-
	43	234.2	26.0	550	1	AV756491	AV756491 AV756491
c	44	234.2	26.0	600	9	AQ075668	AQ075668 CIT-HSP-2
c	45	234.2	26.0	4636	4	BC028737	BC028737 Homo sapi

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS BU785040 507 bp mRNA linear EST 11-OCT-2002
DEFINITION in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3', mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Sceare,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES Location/Qualifiers
source 1. .507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 34.5%; Score 310.4; DB 5; Length 507;
Best Local Similarity 99.7%; Pred. No. 1.8e-37;
Matches 311; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 253

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 193

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 133

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        132 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGCA 73

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 13

Qy        301 CTATCTCAAAAA 312
              ||||||||||||
Db         12 CTATCTCAAAAA 1
```

RESULT 2
CR623485/c
LOCUS CR623485 2213 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DF017YO23 of Fetal brain of Homo sapiens (human).
ACCESSION CR623485
VERSION CR623485.1 GI:50504292
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2213)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 2213)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES Location/Qualifiers
source 1. .2213

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YO23"
/tissue_type="Fetal brain"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 31.2%; Score 281; DB 4; Length 2213;
Best Local Similarity 63.0%; Pred. No. 3e-33;
Matches 550; Conservative 0; Mismatches 285; Indels 38; Gaps 6;

```
Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||||||||| |||| |||||||| | ||| ||||||| || |||| | ||||| ||||
Db       1713 GGGCACAGTGGCTCATGCCTGTAATCTCAGCACTTTGGGAGACCAAAGCGGGTGCATCA- 1655

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
              |||||| || |||||||||||||| | | ||| ||| |||||||| ||||||||||||
Db       1654 -TGAGGTCAGGAGTTCGAGACCAGCTTAGACAACATGATGAAACCCCGTCTCTACTAAAA 1596

Qy        150 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 209
              ||| |||||||||| ||  | |||  |||||||||| ||||||||||||||||||||||
Db       1595 ATATAAAAATTAGCCGGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCGGGAGGCTG 1536

Qy        210 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCAC 269
              ||||| |||||| |||||||| ||||| || |||||||| ||||||||||||    ||||
Db       1535 AGGCAGGAGAATCGCTTGAACTTGGGAAGCGGAGGTTGCGGTGAGCCGAGATTGTGCCAC 1476

Qy        270 TGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 329
              || |||| |||||||| |||| |||||||||| ||| ||| ||| ||| ||| ||||
Db       1475 TGAACTCTAGCCTGGGTGACAGAGCGAGACTCTGTCTAAAACAAACAAACAAACAAAAAC 1416

Qy        330 AAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATT 389
              ||  | |  | | |||| ||   |       | ||| |  |  |       |  || |
Db       1415 AAAAAAC--AAACAAACCAACAAACCTCCTCTGCTGGAATTCCGGGTTATCTCTCCAACT 1358

Qy        390 ATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAA 449
               ||    |   |     |  | |||||||         |   ||  ||||   | ||
Db       1357 TCATCAGTGACTTTCGGTGTAAAGTTTAACTTGGCCTCAGTATCTTTCTTTGTAAAGAAG 1298

Qy        450 AGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGT 509
              || |   |   | |||  |||  |              ||||      |||||    | |
Db       1297 AGAAGGAAGCACCAAAAAAATCTA-------------ATAATATGAGATTATTAGGACT 1252

Qy        510 GGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTT 569
               ||   ||||    ||   | |   |    | | |  | | | |  |   | | ||
Db       1251 CATAGATTGCTATCTATCTTAAATCTGGAAGATAACTGCCAGAGAATTATTTTAAATGTT 1192

Qy        570 GGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTT 629
              ||| | |     |   ||     ||||| ||      | | |      |    | ||
Db       1191 TGGGAAGAAGAAGGAGGTTAGGGGAAGGGAAGAT----GGGAGGAACAGAAGGAGAATTA 1136

Qy        630 TCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTT 689
               | | | ||   || ||| | |  ||     ||||                | | | |
Db       1135 AAAAGAGGCCTTCATAATATACAAGCTACCTCCTTTCC--------------CTTACAT 1091

Qy        690 TAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTGGGA 748
              |||   |     |  ||   |||   |||||||||||||||| ||||||||||||||||
Db       1090 CAAAAGTGGCCTCCGTGGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 1031

Qy        749 GGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATG 808
              ||| ||||    |||||||| ||||||| ||||||  |||||||||||| ||||||  ||
Db       1030 GGCTGAGGAGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATAGTG 971

Qy        809 AAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGT 868
              ||||    |||||| ||||||||||||||||| ||||||   |||||  |  |||||| |
Db       970 AAACCTTGTCTCTACTAAAAATACAAAAATTATCTGGGCGCAGTGGCGTGTGCCTGTAAT 911

Qy        869 CCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              |||||||||||||||  ||||||||||  |||
Db       910 CCCAGCTACTCAGGAGGTTGAGGCAGGAAAATC 878
```

```
RESULT 3
CR626853/c
LOCUS       CR626853               1218 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DF019YF04 of Fetal brain of Homo sapiens
            (human).
ACCESSION   CR626853
VERSION     CR626853.1  GI:50507715
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1218)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1218)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1218
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF019YF04"
                     /tissue_type="Fetal brain"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             31.1%;  Score 280;  DB 4;  Length 1218;
  Best Local Similarity   60.8%;  Pred. No. 5.3e-33;
  Matches  534;  Conservative    0;  Mismatches  325;  Indels   19;  Gaps    4;

Qy         36 AGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGG 95
              |||| |||| |||||| | | ||| ||| ||| || |  ||||||||   |||| ||||
Db       1218 AGTGGCTCATGCCTGTTATCCCAGCACTCTGGGAGGCTGAGGTGGGTAAGTCACCTGAGA 1159

Qy         96 TGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAA 155
              | || ||||| |||||||| ||| ||| ||||  |||||| |||||||||||||| ||||
Db       1158 TCAGGAGTTCAAGACCAGCTTGGCCAACATGGCAAAACCCCATCTCTACTAAAAACACAA 1099

Qy        156 AAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAA 215
              ||| |||||||| || |||  |||  || ||| | | ||||||| ||||||||||||||
Db       1098 AAACTAGCTGGGCGTGGTGGCGCACACCGGTAATTCTAGCTACTTGGGAGGCTGAGGCAT 1039

Qy        216 GAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACT 275
              ||||||  ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1038 GAGAATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCCCACCACTGCACT 979

Qy        276 CCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGAT 335
              |||||||||||||| ||| |||||  |||||||||| | ||  ||||||||      |
Db        978 CCAGCCTGGGCGACAGAGCAAGACTTCATCTCAAAAATAAAATAAAATAAAA------AA 925

Qy        336 CGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGT 395
                |   |  |||| || |  | |   |     |  |||  | |     | || | |
Db        924 TGTTTTGTTTCAAATCTTAATAAAACACACTGTATTTCCAGTAACAAAATCAAAGTAAGA 865

Qy        396 AATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAA 455
                || |  |       |   ||    |      |    | ||||   | | ||  | | |
Db        864 CCAAAGGAATTTTCAGTTGAAAATGCCAGAAGAAAGAATAATTTCTCTGAGAATGGCATA 805

Qy        456 CAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGG 515
              ||  | ||  ||        || |  ||  |   | ||  | |        | ||
Db        804 CAGTCAAATCAAAAGTTTCAGGCAGGCATTAGCTGGCTCTCACTGGTCTCCCTGAGCAGA 745

Qy        516 CATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCA 575
              ||    | || |         |  |   ||    |           || |   |     |
Db        744 CAGACCTGATTCCCACCTGAACTGCCTACATTCTTATCCACAGGGCACCTGCCTTCCTGA 685

Qy        576 TACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAA 635
                |  || ||  | |||          |  | || | |   |   ||   | |||
Db        684 GCCTTTACCTAGCTCTATAAATCTTCTCCTTGGCTTTCATGTAGTACACTTTAAGCAAAA 625

Qy        636 TGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTT----- 690
              ||  || |    |  || | | ||    |    ||||||       ||||| |
Db        624 GGTAAATATCCCTTAACAGTGATGGCATATAACCAATTTTGTTGTAGGATGCTGTTTTCT 565

Qy        691 ------AAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTT 743
                    ||  |   ||  | ||  ||||| ||||| ||| |||||| ||||||||||
Db        564 TTCCCATAAAAACCAAATGTTGGCCGGGCACGGTGGTTCATGCCTGTAATCCCAGCACTT 505

Qy        744 TGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACA 803
              ||||||| ||| |     ||||||| ||||||| |||||| ||||||||||  |||||
Db        504 TGGGAGGACGAAGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCAACA 445

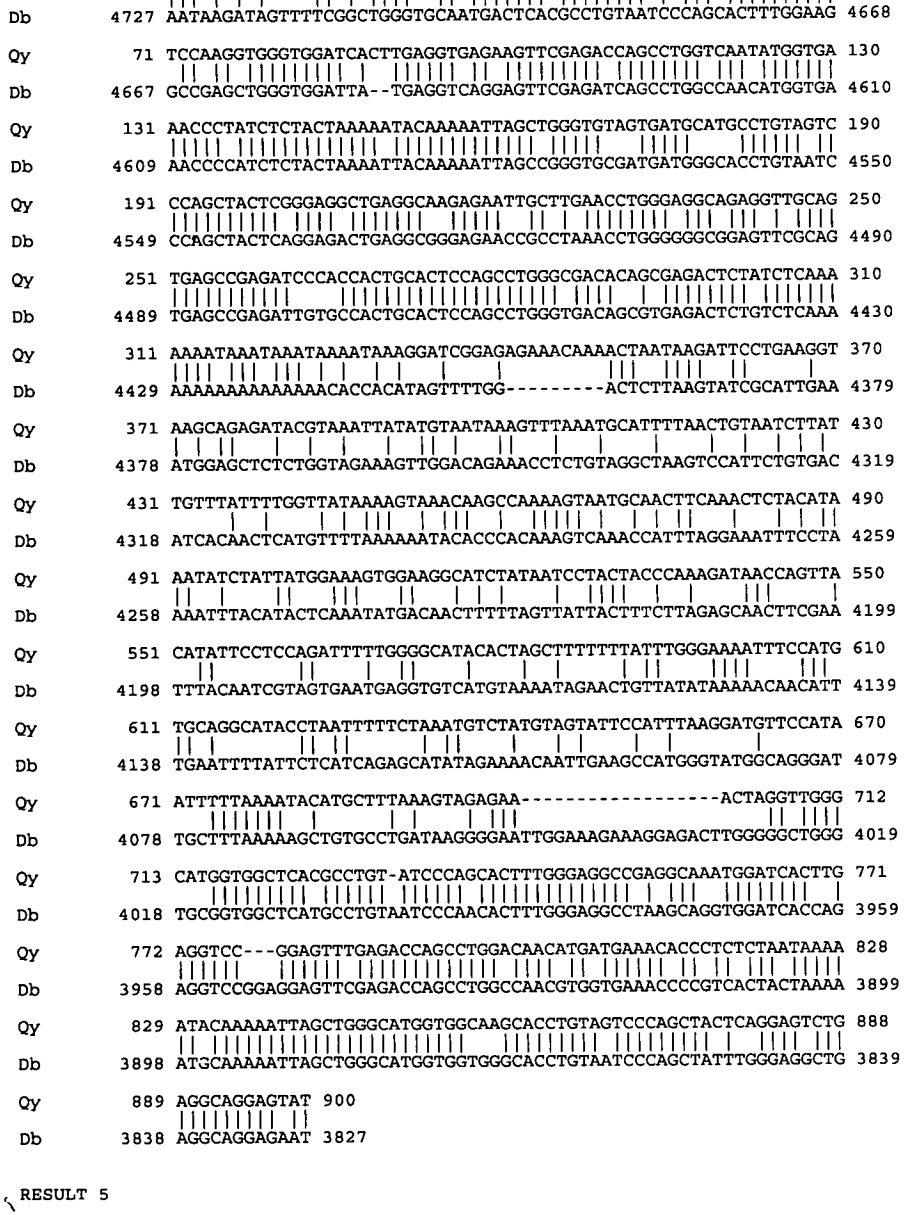
Qy        804 TGATGAAACACCCTCTCTAATAAAAATAC-AAAAATTAGCTGGGCATGGTGGCAAGCACC 862
              |   ||||    |||||| |||||||| ||||||||||||||||||||    || ||
Db        444 CGGCAAAACCTTGTCTCTACCAAAAATACAAAAAATTAGCTGGGCATGGTGATGCGCGCC 385

Qy        863 TGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              |||| |||||||||||  |||| ||||||||||||  ||
Db        384 TGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAAT 347


RESULT 4
CR627178/c
LOCUS       CR627178               6056 bp    mRNA    linear   HTC 22-SEP-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686B03112 (from clone DKFZp686B03112).
ACCESSION   CR627178
VERSION     CR627178.1  GI:50949761
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 6056)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp686B03112) is available at
            the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
            Berlin, Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686B03112
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .6056
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B03112"
                     /tissue_type="fetal kidney"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="fetal"
```

/note="3' UTR"
ORIGIN

Query Match 30.6%; Score 276; DB 4; Length 6056;
Best Local Similarity 60.7%; Pred. No. 1.2e-32;
Matches 554; Conservative 0; Mismatches 325; Indels 33; Gaps 5;

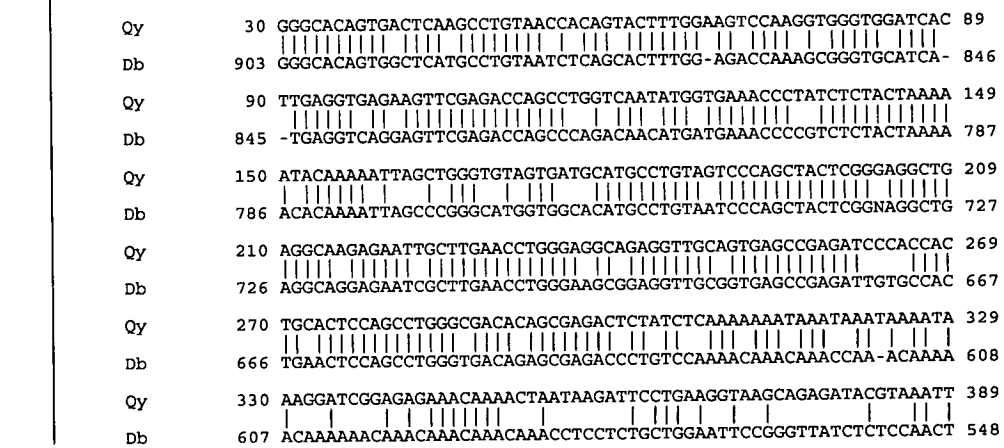


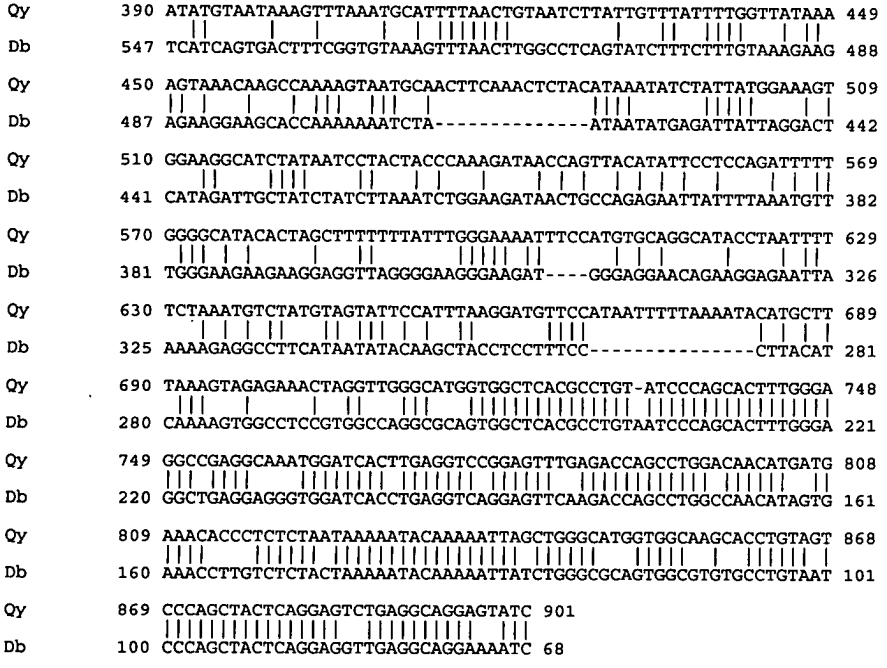
RESULT 5

BX431110/c
LOCUS BX431110 903 bp mRNA linear EST 05-MAY-2004
DEFINITION BX431110 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF017YO23 5-PRIME, mRNA sequence.
ACCESSION BX431110
VERSION BX431110.2 GI:47038087
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 903)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31018008.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.
This sequence belongs to sequence cluster 8501.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BAF014ZE10_AF01326_2&c=8501.r
.

FEATURES Location/Qualifiers
source 1. .903
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YO23"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
ORIGIN

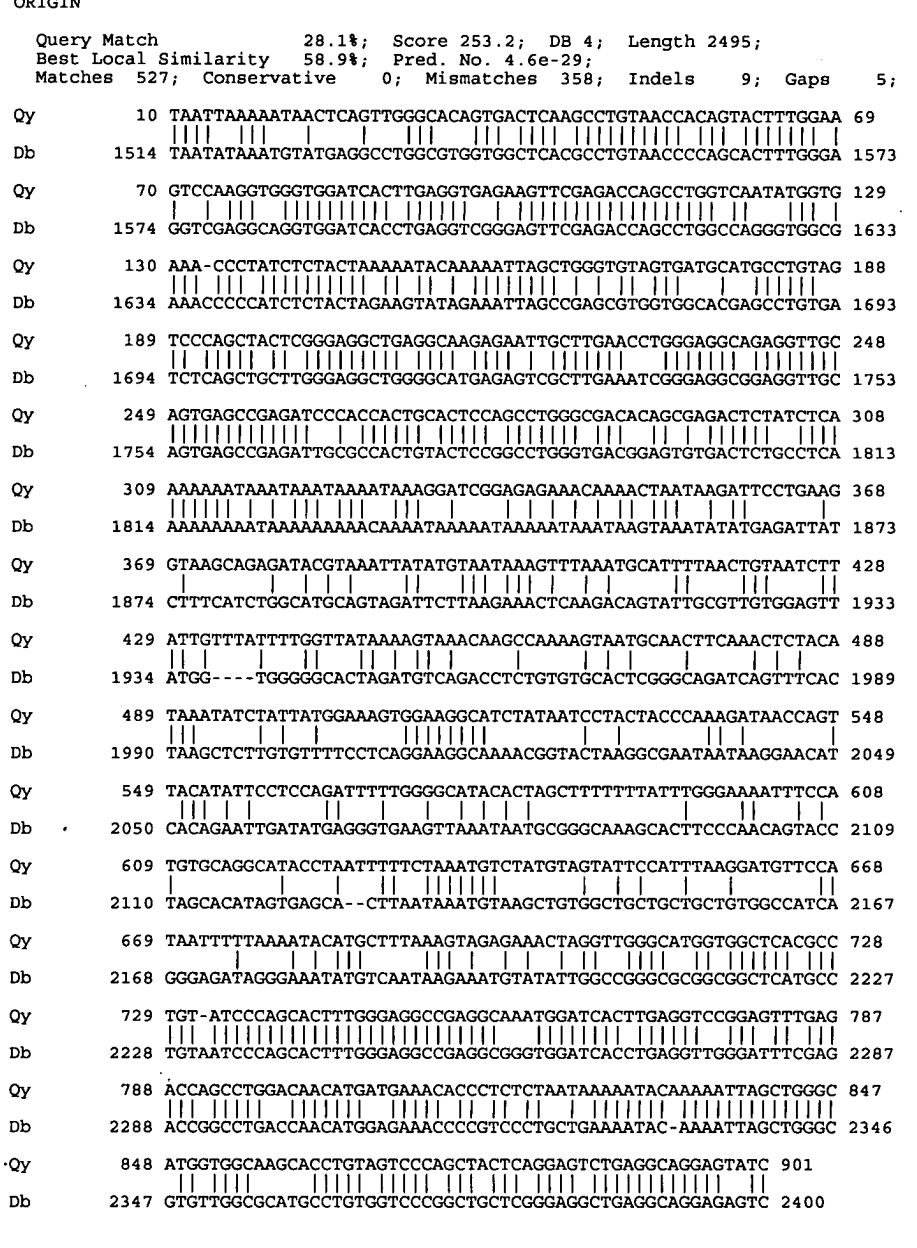
Query Match 28.9%; Score 260.4; DB 5; Length 903;
Best Local Similarity 62.2%; Pred. No. 5.4e-30;
Matches 543; Conservative 0; Mismatches 292; Indels 38; Gaps 7;





```
RESULT 6
HSM808204
LOCUS       HSM808204               2495 bp    mRNA    linear   HTC 20-JAN-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686P04203 (from clone DKFZp686P04203).
ACCESSION   BX648058
VERSION     BX648058.1  GI:34367217
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2495)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2005) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp686P04203) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686P04203
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .2495
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp686P04203Q"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686P04203"
                     /tissue_type="fetal brain"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="fetal"
                     /note="putative transcript"
ORIGIN

  Query Match             28.1%;  Score 253.2;  DB 4;  Length 2495;
  Best Local Similarity   58.9%;  Pred. No. 4.6e-29;
  Matches  527;  Conservative    0;  Mismatches  358;  Indels    9;  Gaps    5;
```



```
RESULT 7
CR614786/c
LOCUS       CR614786               2097 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DF006YD23 of Fetal brain of Homo sapiens
            (human).
ACCESSION   CR614786
VERSION     CR614786.1  GI:50495593
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 2097)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 2097)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .2097
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF006YD23"
                     /tissue_type="Fetal brain"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             28.1%;  Score 253;  DB 4;  Length 2097;
  Best Local Similarity   61.1%;  Pred. No. 5.3e-29;
  Matches  542;  Conservative    0;  Mismatches  290;  Indels   55;  Gaps    6;

Qy         16 AAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAA 75
Db       1329 AATCAAACATGGCTGGGCATGGTGGTTCACGTTTGTAATCCCAGCACTTTGGGAGGCTGA 1270

Qy         76 GGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCC 135
Db       1269 GGTGGGCAGATCAGTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGATGAAACCC 1210

Qy        136 TATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGC 195
Db       1209 CGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGTGGGCACCTGTAATCCCAGC 1150

Qy        196 TACTCGGGAGGC-TGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAG 254
Db       1149 TACTTGGGAGGCCTGAGGCTGAAGAATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAG 1090

Qy        255 CCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAA 314
Db       1089 CCAAGATCACGCCACTGCATTCCAGCCTGGGCAACAGAGCAAGACTCCATCTC-AAAAAA 1031

Qy        315 TAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGC 374
Db       1030 GAAATAAAATAAGGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGCACT-TTGGGAGGC 972

Qy        375 AGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTT 434
Db       971 CGGGGTGGGTGAATCACCTGAGGTCAGGCGTTCAAGACCAGCCTGTCCAACATGGTGAAA 912

Qy        435 TATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATA 494
Db       911 CCCTGTCTCTACTACA----------------------------------AATACAAAAA 886

Qy        495 TCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATA 554
Db       885 TTAGGTGGGCCTGGTGGTGGGTGCCTGTAATCCCAACTACTCAGGAGGCTGAGGCAGGAG 826

Qy        555 TTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCA 614
Db       825 AATCGCTTGAATCCAGGAG----------------TGAGAGGTAGCAGTGCGTCAAGAT 783

Qy        615 GGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTT 674
Db       782 TGCACCACTGCACTCCAGGCTGGACAACAGAGCGAGACTCCGTCTCAAAAAAAAAAAGAA 723

Qy        675 TTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-AT 733
Db       722 AATAATTAAAGAAGTTTAAAAAGAGAGAGACCTCCGGGCATGGTGGCTCATACCTCTAAT 663

Qy        734 CCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGC 793
Db       662 CCCAGCGCTTCGGGAGGAGAAGGCAGGTGGATCATTTGAGGTCAGGAGTTCAAGACCAGC 603

Qy        794 CTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTG 853
Db       602 CTGACCAACATGGTGAAACTCCGTCTCTATTAAAAATACAAAAATTAGCCAGGCATGGTG 543

Qy        854 GCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
Db       542 GCATGCGCCTATAATCCCAGCTACTCAGGAAGCTGAGGCAGGAGAAT 496


RESULT 8
CR607080
LOCUS       CR607080               1945 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI007YG08 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR607080
VERSION     CR607080.1  GI:50487887
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1945)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1945
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI007YG08"
                     /tissue_type="Placenta Cot 25-normalized"
```

```
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match           28.0%;  Score 252.4;  DB 4;  Length 1945;
  Best Local Similarity   61.6%;  Pred. No. 6.7e-29;
  Matches  542;  Conservative    0;  Mismatches  296;  Indels   42;  Gaps    7;

Qy         31 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 90
              ||||| | | |||| |||||||| | ||  |||||||| || |||||||||||  ||||||
Db        666 GGCACGGCGGCTCACGCCTGTAATCCCAACACTTTGGGAGGCCAAGGTGGGCAGATCACC 725

Qy         91 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 150
              |||||| || ||||| |||||||||||||||| |||| ||| || ||||||||||||||
Db        726 TGAGGTCAGGAGTTCAAGACCAGCCTGGTCAACATGGCAAAAGCCCATCTCTACTAAAAA 785

Qy        151 --TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
                |||||||||  || || ||| ||   |||||| ||||| ||||||||||||||
Db        786 ATACAAAAAATTAGCCAGGCGTGGTGGTGGGAGCCTGTAATCCCACCTACTCGGGAGGCT 845

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
              |||||| |||||||||| || |||||||||||||||||||||||| ||||||   |||
Db        846 GAGGCAGGAGAATTGCTTCAATCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATCGTGCCA 905

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
              |||||||||||||||||||| ||  ||||||| ||||| ||||| ||| ||| |||
Db        906 TTGCACTCCAGCCTGGGCGACAGAGTAAGACTCTGTCTCAGAAAAAAAAAAAAAAAAGG 965

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
              |||                  || ||   |  |  ||  ||  | | |    |||||
Db        966 AAA--------------------AAAAAAGATGATATAGTCAAAGAAATAACATTAAAT 1004

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
              |   | ||||  |   | ||||  ||   | || ||  ||  |      |       | ||
Db       1005 TTCTTATAATTTAAGATGAATGATTTCAGATTGAAAAGTTTATAAAGGTGCCAAAGAGAA 1064

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
              ||| ||| |  |   |   |  |  ||  ||    |  | ||  |    ||   |||
Db       1065 AAGGAAAGATCC---ACACTAGATACATTAGAATGAAATGTCAAAGTAACATACATAAA- 1120

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
                ||||  ||    ||  || |||   | | | |||  ||||  |||      | |
Db       1121 --GAAGAAATTCTGAAATCTTCTAGAAAGAAAGAACATGTTATCTATGAAGGAACAAGGA 1178

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
              |       | | ||| ||| |  |       |     | || |   |   |    |||| |
Db       1179 TCATATTGATATTAGATTTCTCAACAGCACATGTGCTGCAAGAAGACAAAGGAGTAATAT 1238

Qy        629 TTCTAAATGTCTA-------TGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAAT 681
              |     | |||  |      |  ||  |   |||| |   |      |     | ||
Db       1239 TAAGTATTGTGAAACAAACCTAGAGCTTAGAATTTCATATTTAAAATTCTAAATAAATTA 1298

Qy        682 ACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCA 740
              |||  |       || ||       ||   ||  |||||||||| |||| | |||||||||
Db       1299 TCATTCAAATTGTTATAGTTTGGGGGCCAGGTGTGGTGGCTCATGCCTATAATCCCAGCA 1358

Qy        741 CTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACA 800
              |||||||||  |  ||||      |||||||||||  || ||||||||||||  |||||||  |
Db       1359 CTTTGGGAGACTAAGGCGGGCAGATCACTTGAGCTCAGGAGTTTGAGACGAGCCTGGCAA 1418

Qy        801 ACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCA 860
              |||  |   |||| || ||||     |||||||||||||||||  |   ||||||
Db       1419 ACACGGCAAAACCCCATCTC-----AAAATACAAAAATTAGCTAGATGTGGTGGTGCATG 1473

Qy        861 CCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900
              ||  || ||| ||||||| |||| ||||||| ||| ||
Db       1474 CCCATAATCCTAGCTACTTGGGAGGCTGAGGCACGAGAAT 1513


RESULT 9
CR604283/c
LOCUS       CR604283               1809 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DF012YB15 of Fetal brain of Homo sapiens
            (human).
ACCESSION   CR604283
VERSION     CR604283.1  GI:50485090
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1809)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1809)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1809
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF012YB15"
                     /tissue_type="Fetal brain"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match           27.8%;  Score 250.2;  DB 4;  Length 1809;
  Best Local Similarity   61.9%;  Pred. No. 1.5e-28;
  Matches  550;  Conservative    0;  Mismatches  298;  Indels   41;  Gaps    8;

Qy         52 AACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACC 111
              || |  |  ||||||| || |  |||  | |||||||| |||| | || |||| ||||||
Db        951 AATCTTATAACTTTGGGAGGCTGAGGCAGATGGATCAC-TGAGATCAGGAGTTTGAGACC 893

Qy        112 AGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAA-TTAGCTGGGTGT 170
              ||| ||| ||| ||| |||||||| ||||   |||||||||||||| ||||||||||||
Db        892 AGCTTGGCCAACATGATGAAACCCCATCTTCTCTAAAAATACAAAAAGTTAGCTGGGTGT 833

Qy        171 AGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAAC 230
              |||      |||||||||||||||||| ||  |||||||||||| ||||||   |||||||
Db        832 GGTGGCAGGTGCCTGTAGTCCCAGCTATTCCAGAGGCTGAGGCAGGAGAATCACTTGAAC 773

Qy        231 CTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACA 290
              ||  ||||| |||||||||||||| |||||| | ||| ||||||||||||||||| |||
Db        772 CTCTGAGGCGGAGGTTGCAGTGAGCTGAGATCGCGCCATTGCACTCCAGCCTGGGCAACA 713

Qy        291 CAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAA 350
              || |||||||| ||||||||||  ||   || ||||||||     |         ||  |
Db        712 GAGTGAGACTCTGTCTCAAAAAATAAATATAAAAAAATAAAAAGCTGCTATTCTCCATTA 653

Qy        351 CTAATAAGATTCCTG--AAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAA 408
              ||      |  || |  |||  ||  |  |  |  ||| ||| ||    |
Db        652 ATACCTGCAGGCCCGACAAAGGCAGCCTAGGAATTCTAGATGAATGGAATGAATACAACT 593

Qy        409 TGC---ATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAA---------- 454
                 |||   |  ||    | || |  | ||  ||||     |  |||  ||
Db        592 TGCTGGAAGTTCTTAGGAAACAAAATGACTATTCACAGAACCAAATAAAAGCCTTCCACT 533

Qy        455 ACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAG 514
              | || | |||| ||||   | |   ||   |   | |||| |||       |  | ||
Db        532 AGAAACTAAAAATAATAATAATAATAATAATGGTTTTATATATATGCATACACAAGCAAA 473

Qy        515 GCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTC---------------- 557
```

```
Db        472 GCCAGAAGAGAATAAATAGCAAATGAATGAAAACTAGAAGCAAAAACAAACAGGAAACCA 413

Qy        558 -CTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGG 616

Db        412 TCTCTACATTTTTTTTACTCAATCGACCCTGAATGCTACAGCATTATCCAGGGCCGCAGA 353

Qy        617 CATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTT 676

Db        352 AAACACACATAATACTTTATTCCTGATAAACAATTCGATGTCCTTAAGTCCGCCAATATC 293

Qy        677 AAAATACATGCTTTAAAGTAGAGAA----ACTAGGTTGGGCATGGTGGCTCACGCCTGT- 731

Db        292 ACCGTACATTCTGTGCAATCAAGAAATTTACTCAGCTGGGCACAGTGGCTCACACCTGTA 233

Qy        732 ATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCA 791

Db        232 ATTCCAGCACTTTGGGAGGCCAAGGCGGGTGGATCACCTGAGGTCAGGCATTCGAGAACA 173

Qy        792 GCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGG 851

Db        172 GCCTGGCCCACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCCGGTATGG 113

Qy        852 TGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTAT 900

Db        112 AGGCATGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAAT 64
```

RESULT 10
AQ839834/c

LOCUS	AQ839834 5468 bp DNA linear GSS 31-MAR-2000
DEFINITION	260L13-C52 CITB Homo sapiens genomic clone 260L13, genomic survey sequence.
ACCESSION	AQ839834
VERSION	AQ839834.1 GI:6652466
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 5468)
AUTHORS	Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R., Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A., Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G., Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and Landes,G.M.
TITLE	A 6-Mb high-resolution physical and transcription map encompassing the hereditary prostate cancer 1 (HPC1) region
JOURNAL	Genomics 64 (1), 1-14 (2000)
PUBMED	10708513
COMMENT	Contact: Carpten JD Cancer Genetics Branch National Human Genome Research Institute/National Institutes of Health Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD Tel: 301 435 5626 Fax: 301 435 5465 Email: jdc@nhgri.nih.gov Class: shotgun.

FEATURES	Location/Qualifiers
source	1. .5468 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="260L13" /clone_lib="CITB"

ORIGIN

Query Match 27.4%; Score 247; DB 9; Length 5468;
Best Local Similarity 57.4%; Pred. No. 3e-28;
Matches 514; Conservative 0; Mismatches 360; Indels 21; Gaps 3;

```
Qy         10 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 69

Db       2985 TAAAGGACAAGTAAGGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 2926

Qy         70 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATAT-GGT 128

Db       2925 GGCCAAGGCCAGCGGATCACCTGAGGTTTGAGGTTCGAGACTAGCCCGGCCAACATGGGT 2866

Qy        129 GAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAG 188

Db       2865 GAAACCCTGTTTCTACTAAAAATACAGAAATTAGCCAGGTGTAGTGGTGTGTGCCTGTAG 2806

Qy        189 TCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGC 248

Db       2805 TCCCAGCTGCTTGGGAGACTGAGGCTGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGC 2746

Qy        249 AGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCA 308

Db       2745 AGTGAGCCGAAATCTTGTCACTGCATTCCAGCCTGGGTGACAAAGTGAGACTCCATCATA 2686

Qy        309 AAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAG 368

Db       2685 AATCAGTCAATCATGTAGTTTCAAGTTTGATTTCCTCTTCAACCCATTCTAACATGTCTT 2626

Qy        369 GTAAGCAGAGATACG--TAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATC 426

Db       2625 TTGCCCCCAAATTAGTCTTACTTGATTCTTATGAAGATTACCATCTTGGCTTCTCTTTAG 2566

Qy        427 TTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTA 486

Db       2565 CTCTGTAACCTCTTCTTGTGAATGTTCCCCTTTATTTACTAGCCACACTGACCTGCTTTC 2506

Qy        487 CATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCA 546

Db       2505 TTTCCTTCTCTAATATGTCTATTTTGAAAGTTATTCCTTGGTTTCCCTATCTTACATGGC 2446

Qy        547 GTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTC 606

Db       2445 TCCTCTTCTCCCCCTTGATTGAAGGTCCCTAGTTTAACTCAT------------------ 2404

Qy        607 CATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTC 666

Db       2403 AAAGTATCAGCATTAGTAAGCAGTAATTCTTTCTGAGTGATGTACTTAGTGATCTTTCAT 2344

Qy        667 CATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACG 726

Db       2343 ATGAATATGAAGGTGGCCTGGCGCAGTGGCTCACTCCTCCACTCATTCCAGCTGGGATTT 2284

Qy        727 CCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGA 786

Db       2283 TACAAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAGGTTGGGAGTTCGA 2224

Qy        787 GACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGG 846

Db       2223 GACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATAGAAAAATTCGCCAGG 2164

Qy        847 CATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901

Db       2163 CGTAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGCGGATC 2109
```

RESULT 11
CR749536/c

LOCUS	CR749536 3737 bp mRNA linear HTC 19-AUG-2004
DEFINITION	Homo sapiens mRNA; cDNA DKFZp686L24211 (from clone DKFZp686L24211).
ACCESSION	CR749536
VERSION	CR749536.1 GI:51476749
KEYWORDS	HTC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 3737)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp686L24211) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686L24211
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3737
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686L24211"
                     /tissue_type="rectum tumor"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
                     /note="transmembrane protein dectin-1, differentially
                     spliced, not fully spliced"
     gene            1. .3737
                     /gene="DKFZp686L24211"
     CDS             131. .541
                     /gene="DKFZp686L24211"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH18342.1"
                     /db_xref="GI:51476750"
                     /translation="MEYHPDLENLDEDGYTQLHFDSQSNTRIAVVSEKGSCAASPPWR
                     LIAVILGILCLVILVIAVVLGTMGEYLVEEINNRQFSKQLLCARHSFKGLTCIIYFNL
                     YIIWFYLYWVGAVNAHFTYKGCERLRGYFTFPGL"
ORIGIN

  Query Match             27.2%;  Score 244.8;  DB 4;  Length 3737;
  Best Local Similarity   60.6%;  Pred. No. 7.4e-28;
  Matches  544;  Conservative    0;  Mismatches  337;  Indels   17;  Gaps    8;

Qy          4 TCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 63
              | | ||    || ||    ||  | ||||  ||||| |||| |||||||| | ||| ||
Db       3229 TTACATATTATACAAGATTCTAGGCTGGGTGCAGTGGCTCACGCCTGTAATCCCAGCAC- 3171

Qy         64 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 123
              |||| || || |||  ||  |||||||||||||| || |||||| ||| | ||||  | |
Db       3170 TTGGGAGGCCGAGGCAGGCAGATCACTTGAGGTCAGGAGTTCGGGACTACCCTGACCGAC 3111

Qy        124 ATGGTGAAACCCTATCTCTACTAAAAATACA-AAAATTAGCTGGGTGTAGTGATGCATGC 182
              |||| |||||||| |||||||||||||| | | ||||||||| |||||| ||| ||| ||
Db       3110 ATGGAGAAACCCTGTCTCTACTAAAAAAATACAAAATTAGCCGGGTGTGGTGGTGCCTGA 3051

Qy        183 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 242
              ||||| |||||||||| |||||||||||||| ||||||||||||||||  || ||| ||
Db       3050 CTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGGGGCGGA 2991

Qy        243 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGAC-ACAGCGAGACTC 301
              |||||| |||||||||| |    ||  ||||||||||||||||| || | ||||| ||||
Db       2990 GGTTGCGGTGAGCCGAGACCGTGCCGTTGCACTCCAGCCTGGGCAACAAGAGCGAAACTC 2931

Qy        302 TATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATT 361
              | |||||||||||      ||| |||| ||||  ||    ||| | |   |  ||    |
Db       2930 TGTCTCAAAAAA------AAAAAAAAAAAAAGATTCTATTAGAGATATGGCAGATGTACTC 2877

Qy        362 CCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTG 421
              ||||        |  ||||   |||   || |   |   | ||      |   | |
Db       2876 TCTGAGCTATTTAATTGATAAAGAAAAAGTAGGAGTTCCTGCTTTTTACTAAGATGATAA 2817

Qy        422 TAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAA 481
              | ||  |  || | ||  || | | || | ||||   ||    | |   |  |    |
Db       2816 TCATAATTATGATCATACTGATGAAAATAATAAAATTGCTCTGACTGTGGTGAA--GATT 2759

Qy        482 CTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGAT 541
              | | |        ||  |||   |    | ||   || |       |   |      | |||
Db       2758 TTTTCCCATTTCTTCATTTAGCTATTCAGAAATATATTTCACCCTCAGTTCATAACTGAT 2699

Qy        542 AACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAA 601
              | | | || |  ||        |  |   ||  | ||||       ||  ||    |  |
Db       2698 ATACTGCTAGAAGTTGAGGGTCAAATCGTGGCAACACACCGTGCACTTCAAT----GGCA 2643

Qy        602 ATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGA 661
              | | |     |  ||   || ||    | | |  |  |  |    ||     | |
Db       2642 TTGTTGAGCATGAAACAACTCTGATAGAGCCAGAAATTAAGATTCGCTTATGGCTTACCC 2583

Qy        662 TGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGC 721
                        |||    | | ||| | |        ||   ||  ||||||||
Db       2582 AAGGTAAATGCCAAGTCAATGGCAGAATGAAAATTCAAGCTTGGGGCCAGGTGTGGTGGC 2523

Qy        722 TCACGCCTGT-ATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGA 780
              ||||  |||| ||||||||||||||||||||| ||||      |||||||| ||  | |||
Db       2522 TCACATCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGCAGATCACTTAAGCCCAGGA 2463

Qy        781 GTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATA-CAAAAATT 839
              ||| |||||||| ||| |||||||   |||| |  ||||||  ||||| | ||||||||
Db       2462 GTTAGAGACCAGCTTGGTCAACATGGCAAAACCCAGTCTCTACAAAAAACACCAAAAATT 2403

Qy        840 AGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGAG 897
              |||  |||||||||||| ||||||||||||||||||||||    | | |||||||  |||
Db       2402 AGCCTGGCATGGTGGCATGCACCTGTAGTCCCAGCTACTTGAGGGGCTGAGGCGAGAG 2345

RESULT 12
CN414771
LOCUS       CN414771                669 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000600839160 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN414771
VERSION     CN414771.1  GI:47402365
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
   PUBMED   15146197
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 669   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                     line"
                     /clone_lib="GRN_PREHEP"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from DMSO-treated hES cell line H9 (p22) maintained in
```

feeder-free conditions"
ORIGIN

Query Match 27.1%; Score 244.6; DB 7; Length 669;
Best Local Similarity 83.7%; Pred. No. 1.5e-27;
Matches 277; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Qy          1 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 60
              || ||||  |  | |||| |    |||  |||||||||| |||| ||||||||  | |||
Db         46 TGATCATCCTGTTGAAAATTGCAACAGCCGGGCACAGTGGCTCATGCCTGTAATCCCAGC 105

Qy         61 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 120
              |||||||||| || |||||||||||||| | |||| | || ||||||||||||||||||  |
Db        106 ACTTTGGAAGGCCGAGGTGGGTGGATCGCCTGAGATCAGGAGTTCGAGACCAGCCTGACC 165

Qy        121 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 180
              || |||||||||||||| |||||||||||||||||||||||||||||| || ||| ||  |
Db        166 AACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGTGGGT 225

Qy        181 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 240
              |||||| |||| |||| || |||||||||| || |||||| ||||||||||||||||||
Db        226 GCCTGTGGTCCCGGCTAATCAGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGGAGGCA 285

Qy        241 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 300
              ||||||||||||||||  ||||   ||| ||||||||||||||||  |||| ||| |||||
Db        286 GAGGTTGCAGTGAGCCAGGATCATGCCATTGCACTCCAGCCTGGATGACAGAGCAAGACT 345

Qy        301 CTATCTCAAAAAAATAAATAAATAAAATAAA 331
              || ||||||| |||||||||||||||| |||
Db        346 CTGTCTCAAATAAATAAATAAATAAAAGAAA 376
```

RESULT 13
CD102556/c
LOCUS CD102556 691 bp mRNA linear EST 15-MAY-2003
DEFINITION AGENCOURT_14024251 NIH_MGC_186 Homo sapiens cDNA clone IMAGE:30369934 5', mRNA sequence.
ACCESSION CD102556
VERSION CD102556.1 GI:30755730
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 691)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: NDCM158 row: b column: 23
High quality sequence stop: 509.
FEATURES Location/Qualifiers
source 1. .691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30369934"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_186"
/note="Organ: Pooled-Skin; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); Library is oligo-dT primed and directionally cloned. cDNA was prepared from a pooled samples of tissues from Skin, meninges, duramatter, pia matter and choroid plexus. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library"

ORIGIN

Query Match 27.0%; Score 243.6; DB 6; Length 691;
Best Local Similarity 82.5%; Pred. No. 2.1e-27;
Matches 279; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          5 CATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTT 64
              || | |||| |  || | ||||| |||||||||||| |||| |||||||| | ||| ||||
Db        661 CAAAATAATAAGGAACATCTCAGCTGGGCACAGTGGCTCACGCCTGTAATCCCAGCACTT 602

Qy         65 TGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATA 124
              ||| || ||||||||||  ||||||||||||  || ||||| |||||||||| | ||| |
Db        601 TGGGAGGCCAAGGTGGGCAGATCACTTGAGGCCAGGAGTTCAAGACCAGCCTAGCCAACA 542

Qy        125 TGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCT 184
              |||||||| || |||||||||||||||||||| |||||||||  | ||| ||||  |||
Db        541 TGGTGAAATCCCATCTCTACTAAAAATACAAAACTTAGCTGGGCATGGTGGTGCACCCCT 482

Qy        185 GTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGG 244
              |||||||||||||| | | |||||||||| | ||||||||||||||  ||||||||||||
Db        481 GTAGTCCCAGCTACCCAGTAGGCTGAGGTAGGAGAATTGCTTGAACACGGGAGGCAGAGG 422

Qy        245 TTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTAT 304
              ||||||||||||||||||   |||||||||||||||| || |||| || ||||||||  |
Db        421 TTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTAGGTGACAGAGAGAGACTCTGT 362

Qy        305 CTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAG 342
              ||||   ||| || |||| |||| ||| | | || |||
Db        361 CTCATTGAAAAAATTAAAAAAAAAAAATGCTGGGTGAG 324
```

RESULT 14
AG186254
LOCUS AG186254 710 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-060J11.T7, genomic survey sequence.
ACCESSION AG186254
VERSION AG186254.1 GI:16715934
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 710)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY

```
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES            Location/Qualifiers
     source         1. .710
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /clone="RP43-060J11.T7"
                    /sex="male"
                    /cell_type="lymphocytes"
                    /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN

  Query Match            27.0%;  Score 243.2;  DB 10;  Length 710;
  Best Local Similarity  76.3%;  Pred. No. 2.4e-27;
  Matches  299;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy          6 ATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTT 65
              || | || |||||| ||    | ||| | |  || |||| || ||||| | ||| |||||
Db         58 ATTTAAAATAAAAAAAATAAGGCTGGACGCGATGGCTCACGCTTGTAATCCCAGCACTTT 117

Qy         66 GGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATAT 125
              || || |||||||||||||||||| |||||| || |||| | |||||||||||| ||| ||
Db        118 GGGAGGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGGGACCAGCCTGGCCAACAT 177

Qy        126 GGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTG 185
              || |||||||| |||||||||| | |||||||||| |||| | ||| || ||||| |
Db        178 GGCGAAACCCTGTCTCTACTAAAGAGACAAAAATTAGTTGGGCTTGGTGGTGGATGCCAG 237

Qy        186 TAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGT 245
              || |||||||||||| |||||||||||||| |||||||||||||||| ||||||||| ||
Db        238 TAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCAGAAGT 297

Qy        246 TGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATC 305
              |||||||||||||||| |  | |||||||||||||||||| ||| || |||||||| ||
Db        298 TGCAGTGAGCCGAGATCACGACGCTGCACTCCAGCCTGGGCAACAGAGTGAGACTCTGTC 357

Qy        306 TCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTG 365
              ||||||| | |||  |||||||||||  |    ||| ||      || | ||  ||| |
Db        358 TCAAAAATAAAAAATAATAAAATAAAATAAATTAGAAAATATCTTCTCAAAAATTTCATA 417

Qy        366 AAGGTAAGCAGAGATACGTAAATTATATGTAA 397
              | |    |    |  |||||||| |  ||
Db        418 AGAGGTGAAAAGTTGAATTAAATTATTTTAAA 449


RESULT 15
AF370428/c
LOCUS       AF370428               4640 bp    mRNA    linear   HTC 31-DEC-2003
DEFINITION  Homo sapiens PP7302 mRNA, complete cds.
ACCESSION   AF370428
VERSION     AF370428.1  GI:33341785
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 4640)
  AUTHORS   Huang,Y., Zhou,X.M., Zhang,P.P., Jiang,H.Q., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Novel human cDNA clones with function of inhibiting cancer cell
            growth
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4640)
  AUTHORS   Zhou,X.M., Zhang,P.P., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T.,
            Wan,D.F. and Gu,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2001) National Laboratory For Oncogenes & Related
            Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
            200032, P. R. China
FEATURES            Location/Qualifiers
     source         1. .4640
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
     CDS            2431. .3039
                    /codon_start=1
                    /product="PP7302"
                    /protein_id="AAQ15264.1"
                    /db_xref="GI:33341786"
                    /translation="MGKGKRETKNIGKRTDETTKEEPIPNERGLGQATCTLFQILTQR
                    YLIASGQCFTVYKGLCYQDNPTVTSILDARSTKSGSRLHSKYTRSRDLNLFSVFLQCY
                    YFIPEFPLPSLCLLSLLCPLPTQKVHTNLLCSCARGLVGEWRWTSNTACSRWKVAVPA
                    HTRYTAPEPLSTGQSERPNRWAHNPEGIWIWECVCEKWQDKS"
ORIGIN

  Query Match            26.9%;  Score 242;  DB 4;  Length 4640;
  Best Local Similarity  60.6%;  Pred. No. 1.8e-27;
  Matches  531;  Conservative    0;  Mismatches  315;  Indels   30;  Gaps    7;

Qy         30 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 89
              |||  ||||| |||| |||||||| | ||| ||||||| || |  ||| ||| |||||||
Db       2206 GGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCGGATCAC 2147

Qy         90 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 149
                ||||| |  || || |||| | |||||  || ||||||||||||| ||||||||||||
Db       2146 --GAGGTCAAGAGATCAAGACAATCCTGGCTAACATGGTGAAACCCCATCTCTACTAAAA 2089

Qy        150 ATACAAAA-ATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
              |||||||| |||||| ||| | ||| |  |||||| |||||||| ||| |||||
Db       2088 ATACAAAATATTAGCCGGGCATGGTGGCACGCACCTGTAATCCCAGCTGCTCAGGAGGTG 2029

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
              ||||||  |||||  |||||||| ||||||||||||||||||||||| || ||| |||||
Db       2028 GAGGCAGAAGAATCACTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAAATCACACCA 1969

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
              |||||||| |||||||| |||| ||| ||||||| |||||||||| ||| || ||| |
Db       1968 CTGCACTCTAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAGTAACAA 1909

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAAT 388
                |           | |  |  | ||    ||    |  || ||  |
Db       1908 TTTGACATTTTTCCACTGGTGCCCATCTTGCAGCATGGCATTACATTGGCATCTGTGGAG 1849

Qy        389 TATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAA 448
                | || | | |   |                  ||    |||  |  |  ||| |
Db       1848 GCTTTGCATTCATCCT--------------------ACTGCAGGTTGGTGCTTATTACAG 1809

Qy        449 AAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAG 508
                |  | |  |       | | |     ||     ||||||   |  |  |
Db       1808 CTTTTGCCCATTCCTGGAACAAGAACTGGTAAGGGGCACATAACCCTAAAGGCTTTTCCT 1749

Qy        509 TGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTT 568
              || |||  ||||  ||  ||||||     | |    |  |  | |  |    |  ||||
Db       1748 GGGGAGGTTTCTAGTAT--TACTACATGGAAAGGGCTGGGCAAAAGGCACGGCAGAGCAC 1691

Qy        569 TGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTT 628
              |||  |    ||  |   |      |  ||||   |   |    |  ||  ||  |
Db       1690 TGGTCCCACAGCTTCCCACTGAGCAGTATCAAAAAAGCTTAGGAATGTGATGACTGCGTG 1631

Qy        629 TTCTAAATGTCTATG--TAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATG 686
                | || |      |       | ||| |      |  |       |     ||   |
Db       1630 GGCCAAGTAGGATGGCTAGTCACACCAGTCTTTTGAGAGCAGGGGAGTAGTTGGTAATGG 1571

Qy        687 CTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGT-ATCCCAGCACTTTG 745
              |  ||       ||      | |||||| |||||||| |||||| ||||||||||||||
Db       1570 GTAAAAGAGTTGGAGCTCCAGCTGGGCACGGTGGCTCATGCCTGTAATCCCAGCACTTTG 1511

Qy        746 GGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATG 805
              ||||||||||    |||||||  ||||||| |||||||||||||||||||||| |||||||
Db       1510 GGAGGCCGAGGTGGGTGGATCA--TGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATG 1453

Qy        806 ATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGT 865
              |||||| |||||||| |||||||| |||| |||||||| |||||||| | |  |||||
```

```
Db       1452 GTGAAACCCCCTCTCTAGTAAAAATATAAAAGTTAGCTGGACATGGTGGTAGGTGCCTGT 1393

Qy        866 AGTCCCAGCTACTCAGGAGTCTGAGGCAGGAGTATC 901
              | ||||||| ||||||||| |||||| ||||| |||
Db       1392 AATCCCAGCCACTCAGGAGGCTGAGGTAGGAGAATC 1357
```

Search completed: February 10, 2006, 00:58:12
Job time : 3374.06 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 3492.02 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 699.8
Sequence: 1 gtcgtgtcactgctgtccag.....tgcagcctcaacctcctgag 701

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	699.8	100.0	8368	8	AY438977	AY438977 Homo sapi
c	2	699.8	100.0	122302	8	AC003982	AC003982 Homo sapi
c	3	699.8	100.0	220384	14	AC078926	AC078926 Homo sapi
	4	699.4	100.0	13612	6	AX377239	AX377239 Sequence
	5	380.8	54.4	3375	8	HUMPLA2A1	M22970 Human pancr
	6	355.8	50.8	357	10	G67590	G67590 csnppla2g1b
	7	312	44.6	189729	14	AC073930	AC073930 Homo sapi
c	8	211.4	30.2	227930	14	AC158058	AC158058 Bos tauru
c	9	140.6	20.1	215223	14	AC127942	AC127942 Rattus no
	10	140.6	20.1	238961	14	AC097575	AC097575 Rattus no
c	11	140.6	20.1	290053	14	AC096966	AC096966 Rattus no
	12	130.8	18.7	418	6	CQ922146	CQ922146 Sequence
	13	130.8	18.7	447	6	AX377240	AX377240 Sequence
	14	130.8	18.7	559	6	CQ723247	CQ723247 Sequence
	15	130.8	18.7	562	6	AR380655	AR380655 Sequence
	16	130.8	18.7	562	6	AX332734	AX332734 Sequence
	17	130.8	18.7	562	6	AX333340	AX333340 Sequence
	18	130.8	18.7	562	6	AX333743	AX333743 Sequence
	19	130.8	18.7	562	8	HUMPLA2RA	M21054 Human lung
	20	130.8	18.7	569	6	E02268	E02268 cDNA encodi
	21	130.8	18.7	644	6	BD209899	BD209899 Human nuc
	22	130.8	18.7	644	6	AX014213	AX014213 Sequence
	23	130.8	18.7	654	6	CS071209	CS071209 Sequence
c	24	126.2	18.0	496	6	CS071001	CS071001 Sequence
	25	126	18.0	10234	9	AF094611	AF094611 Mus muscu
c	26	126	18.0	183861	9	AC117735	AC117735 Mus muscu
	27	126	18.0	215516	14	AC024694	AC024694 Mus muscu
	28	126	18.0	238599	9	AC159539	AC159539 Mus muscu
	29	96.8	13.8	562	4	DOGPHA2	M35301 Canine phos
	30	96.8	13.8	567	4	DOGPPLA2	D00035 Canis sp. m
c	31	95.2	13.6	154362	14	AC069517	AC069517 Homo sapi
	32	95.2	13.6	228098	8	AL596247	AL596247 Human DNA
	33	92	13.1	542	9	RATPPLA2	D00036 Rattus norv
	34	90.8	13.0	546	4	RABPHOPHO	L11899 Oryctolagus
	35	90.4	12.9	527	9	AF187852	AF187852 Mus muscu
	36	90.4	12.9	531	9	AF097637	AF097637 Mus muscu
	37	90.4	12.9	542	9	AF094610	AF094610 Mus muscu
	38	90.4	12.9	552	9	AF162712	AF162712 Mus muscu
	39	89.8	12.8	369	6	E01687	E01687 cDNA encodi
	40	89.8	12.8	546	4	BTPLA2R	Y00120 Bos taurus
	41	89	12.7	173929	14	AC131928	AC131928 Homo sapi
c	42	89	12.7	189230	8	AC103974	AC103974 Homo sapi
c	43	89	12.7	198565	14	AC016750	AC016750 Homo sapi
	44	88.2	12.6	483	4	SSPLA2R	Y00146 Porcine mRN
	45	88.2	12.6	509	4	PIGPLA2RA	M21055 Pig pancrea

ALIGNMENTS

RESULT 1
AY438977
LOCUS AY438977 8368 bp DNA linear PRI 29-OCT-2003
DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
ACCESSION AY438977
VERSION AY438977.1 GI:37953284
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8368)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).
FEATURES Location/Qualifiers
source 1. .8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101. .257
/rpt_family="MIR"
/rpt_type=dispersed
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326. .538
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 354. .496

```
                /note="Region not scanned for variation"
gene            875. .6548
                /gene="PLA2G1B"
mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                /gene="PLA2G1B"
                /product="phospholipase A2, group IB (pancreas)"
CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                /gene="PLA2G1B"
                /codon_start=1
                /product="phospholipase A2, group IB (pancreas)"
                /protein_id="AAR05441.1"
                /db_xref="GI:37953285"
                /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
repeat_region   1174. .1307
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1286
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="g"
repeat_region   1319. .1588
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1343
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="a"
variation       1593
                /gene="PLA2G1B"
                /frequency="0.02"
                /replace="t"
repeat_region   1605. .1669
                /rpt_family="L2"
                /rpt_type=dispersed
misc_feature    1638. .2188
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
repeat_region   1670. .1975
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   1976. .2038
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2086. .2165
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2194. .2319
                /rpt_family="Alu"
                /rpt_type=dispersed
misc_feature    2319. .2453
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
variation       3027
                /gene="PLA2G1B"
                /frequency="0.25"
                /replace="t"
repeat_region   3165. .3431
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       3262
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3321
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
variation       3396
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="g"
variation       3630
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="c"
variation       3702
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="a"
variation       3768
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3818
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
repeat_region   3846. .3964
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   3977. .4280
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   4281. .4346
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   4347. .4634
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       4386
                /gene="PLA2G1B"
                /frequency="0.08"
                /replace="t"
repeat_region   4635. .4723
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5013
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="a"
repeat_region   5103. .5404
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5133
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5543
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5555
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       5556
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="g"
repeat_region   5577. .5629
                /rpt_family="L1"
                /rpt_type=dispersed
variation       5580
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
variation       5675
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
repeat_region   5779. .6083
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5929
```

```
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             100.0%;  Score 699.8;  DB 8;  Length 8368;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-179;
  Matches  698;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3374 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3433

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3434 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3493

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3494 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3553

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3554 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3613

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3614 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3673

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3674 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3733

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3734 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 3793

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3794 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 3853

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3854 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 3913

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3914 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 3973

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3974 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 4033

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db       4034 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 4074


RESULT 2
AC003982/c
LOCUS       AC003982              122302 bp    DNA     linear   PRI 18-MAR-1999
DEFINITION  Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION   AC003982
VERSION     AC003982.1  GI:2769695
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122302)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 122302)
  AUTHORS   Bradshaw,H., Wu,X. and Ozersky,P.
  TITLE     The sequence of Homo sapiens PAC clone 166H1
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was originally isolated in the laboratory of Professor
            Graeme Bell, Howard Hughes Medical Institute and Departments of
            Biochemistry and Molecular Biology, and Medicine, The University of
            Chicago, Chicago, IL, USA.  The clone was provided by the
            laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
            Genetics, Oxford, UK.  Some contig information was also obtained
            from Yamagata et al., Nature 384:455-8 (1996).

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is 278C19; the clone sequenced to
            the right is 15E1.  Actual start of this clone is at base position
            1 of 166H1; actual end is at 122302 of 166H1.
FEATURES                Location/Qualifiers
     source             1. .122302
```

```
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q"
                     /clone="166H1"
                     /clone_lib="RPCI-1"
     repeat_region   1. .238
                     /rpt_family="Alu"
     repeat_region   239. .265
                     /rpt_family="AT_rich"
     repeat_region   268. .565
                     /rpt_family="Alu"
     repeat_region   1001. .1309
                     /rpt_family="Alu"
     repeat_region   1440. .1745
                     /rpt_family="Alu"
     repeat_region   1909. .1940
                     /rpt_family="MER1_type"
     repeat_region   1955. .1983
                     /rpt_family="AT_rich"
     repeat_region   2063. .2362
                     /rpt_family="Alu"
     repeat_region   2516. .2650
                     /rpt_family="Alu"
     repeat_region   2674. .2730
                     /rpt_family="L1"
     repeat_region   2735. .2788
                     /rpt_family="(GAAAA)n"
     repeat_region   2790. .3077
                     /rpt_family="Alu"
     repeat_region   3081. .3139
                     /rpt_family="L1"
     repeat_region   3716. .4020
                     /rpt_family="Alu"
     repeat_region   4196. .4492
                     /rpt_family="Alu"
     gene            4560. .14034
                     /gene="WUGSC:H_166H1.1"
     CDS             join(4560. .5056,13418. .13748,13882. .14034)
                     /gene="WUGSC:H_166H1.1"
                     /note="unknown function; 60% similar to Z50177
                     (PID:g927403) (PID:g927402); H_166H1.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB95634.1"
                     /db_xref="GI:2769696"
                     /translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
                     KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
                     QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
                     HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
                     LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
                     VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
     repeat_region   5366. .5655
                     /rpt_family="Alu"
     repeat_region   5672. .5798
                     /rpt_family="Alu"
     repeat_region   5802. .5834
                     /rpt_family="(TA)n"
     repeat_region   5834. .5857
                     /rpt_family="POLY_A"
     repeat_region   5870. .6184
                     /rpt_family="Alu"
     repeat_region   6224. .6268
                     /rpt_family="(CA)n"
     repeat_region   6269. .6558
                     /rpt_family="Alu"
     repeat_region   6729. .7029
                     /rpt_family="Alu"
     repeat_region   7051. .7172
                     /rpt_family="Alu"
     repeat_region   7192. .7500
                     /rpt_family="Alu"
     repeat_region   7518. .7820
                     /rpt_family="Alu"
     repeat_region   8207. .8500
                     /rpt_family="Alu"
     repeat_region   8576. .8878
                     /rpt_family="Alu"
     repeat_region   9023. .9154
                     /rpt_family="Alu"
     repeat_region   9165. .9470
                     /rpt_family="Alu"
     repeat_region   9887. .10188
                     /rpt_family="Alu"
     repeat_region   10221. .10342
                     /rpt_family="L1"
     repeat_region   10408. .10707
                     /rpt_family="Alu"
     repeat_region   10713. .11014
                     /rpt_family="Alu"
     repeat_region   11020. .11110
                     /rpt_family="Alu"
     repeat_region   11146. .11386
                     /rpt_family="Alu"
     repeat_region   11208. .11386
                     /rpt_family="Alu"
     repeat_region   11391. .11515
                     /rpt_family="Alu"
     repeat_region   11568. .11705
                     /rpt_family="Alu"
     repeat_region   11737. .12032
                     /rpt_family="Alu"
     repeat_region   12034. .12066
                     /rpt_family="7SLRNA"
     repeat_region   12075. .12372
                     /rpt_family="Alu"
     repeat_region   12448. .12753
                     /rpt_family="Alu"
     repeat_region   12763. .12874
                     /rpt_family="L2"
     repeat_region   12999. .13303
                     /rpt_family="Alu"
     misc_feature    complement(13674. .13752)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     misc_feature    complement(13840. .14195)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     repeat_region   14196. .14229
                     /rpt_family="AT_rich"
     repeat_region   14240. .14339
                     /rpt_family="U6"
     repeat_region   14349. .14646
                     /rpt_family="Alu"
     repeat_region   14652. .14940
                     /rpt_family="Alu"

  Query Match            100.0%;  Score 699.8;  DB 8;  Length 122302;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-179;
  Matches  698;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26288 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 26229

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26228 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 26169

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26168 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 26109

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26108 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 26049
```

```
Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db      26048 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 25989

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25988 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 25929

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25928 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 25869

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25868 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 25809

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25808 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 25749

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25748 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 25689

Qy        601 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25688 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 25629

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db      25628 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 25588
```

```
RESULT 3
AC078926/c
LOCUS       AC078926               220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
```

```
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match             100.0%;  Score 699.8;  DB 14;  Length 220384;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-179;
  Matches  698;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71511 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 71452

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71451 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 71392

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71391 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 71332

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71331 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 71272

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db      71271 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 71212

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71211 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 71152

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71151 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 71092

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71091 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 71032

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71031 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 70972

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70971 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 70912

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70911 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 70852

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db      70851 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 70811


RESULT 4
AX377239
LOCUS       AX377239               13612 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212562.
ACCESSION   AX377239
VERSION     AX377239.1  GI:19573528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 1 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .13612
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     variation       3845
                     /note="PS1: polymorphic base G or A"
     variation       3968
                     /note="PS2: polymorphic base A or G"
     variation       6060
                     /note="PS2: polymorphic base A or G"
     variation       6844
                     /note="PS4: polymorphic base G or A"
     variation       9531
                     /note="PS5: polymorphic base G or A"
ORIGIN

  Query Match             100.0%;  Score 699.8;  DB 6;  Length 13612;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-179;
  Matches  697;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815
```

```
Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::|||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       6816 ACCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7116 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db       7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 7216
```

```
RESULT 5
HUMPLA2A1
LOCUS       HUMPLA2A1               3375 bp    DNA     linear   PRI 07-JAN-1995
DEFINITION  Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.
ACCESSION   M22970 M14965
VERSION     M22970.1  GI:190008
KEYWORDS    phospholipase A2.
SEGMENT     1 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3375)
  AUTHORS   Seilhamer,J.J., Randall,T.L., Yamanaka,M. and Johnson,L.K.
  TITLE     Pancreatic phospholipase A2: isolation of the human gene and cDNAs
            from porcine pancreas and human lung
  JOURNAL   DNA 5 (6), 519-527 (1986)
   PUBMED   3028739
COMMENT     Original source text: Homo sapiens pulverized liver DNA.
FEATURES             Location/Qualifiers
     source          1. .3375
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="12q23-qter"
                     /tissue_type="pulverized liver"
     exon            <537. .570
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=1
     intron          571. .2262
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=1
     exon            2263. .2422
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=2
     intron          2423. .3222
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=2
     exon            3223. .3350
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=3
     intron          3351. .>3375
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=3
ORIGIN      5 bp upstream of PstI site.
```

```
  Query Match             54.4%;  Score 380.8;  DB 8;  Length 3375;
  Best Local Similarity   99.2%;  Pred. No. 1e-92;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3053

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3113

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3173

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3233

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       3234 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3293

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3353

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||||
Db       3354 GGTTTATCCCTTCCTTGACCTA 3375
```

```
RESULT 6
G67590
LOCUS       G67590                  357 bp    DNA     linear   STS 31-JAN-2001
DEFINITION  csnppla2g1b-pcr3-1 Human Homo sapiens STS genomic, sequence tagged
            site.
ACCESSION   G67590
VERSION     G67590.1  GI:12642952
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Wong,G.K.S., Yu,J., Yang,Z., Passey,D., Kibukawa,M., Paddock,M. and
            Olson,M.
  TITLE     Gene based polymorphism discovery
  JOURNAL   Unpublished (2000)
COMMENT
            Contact: Gane Ka-Shu Wong
            Genome Center
            University of Washington
            Fluke Hall, Box 352145, Seattle, WA 98195, USA
            Tel: 206/685-7348
            Fax: 206/685-7344
            Email: gksw@u.washington.edu
            Primer A: AAGCATCTTGGCGAAGTCAT
            Primer B: AAAGGAGACACTGCCCAGAA
            STS size: 357
            PCR Profile:
```

```
                Presoak:92 degrees C for 1.00 minute
                Denaturation:92 degrees C for 0.16 minute
                Annealing:60-65 degrees C for 0.50 minute
                Polymerization:72 degrees C for 1.00 minute
                PCR Cycles:35
                Thermal Cycler:Perkin Elmer TC
           Protocol:
                Template:     7 ng
                Primer:     each 0.5 uM
                dNTPs:     each 100 uM
                Taq Polymerase:     0.025 units/ul
                Total Vol:     10 ul

           Buffer:
                MgCl2:     1.5 mM
                KCl:     50 mM
                Tris-HCl:     10 mM
                pH:     8.3.
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="12q23-q24.1"
                     /clone_lib="Human"
     gene            1. .357
                     /gene="PLA2G1B"
     STS             1. .357
                     /gene="PLA2G1B"
     primer_bind     1. .20
                     /gene="PLA2G1B"
     primer_bind     complement(338. .357)
ORIGIN

  Query Match             50.8%;  Score 355.8;  DB 10;  Length 357;
  Best Local Similarity   99.2%;  Pred. No. 7e-86;
  Matches  354;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        123 AAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGACCT 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGACCT 60

Qy        183 ATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACA 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACA 120

Qy        243 CATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMM 302
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||::
Db        121 CATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAAC 180

Qy        303 CCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGG 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGG 240

Qy        363 TTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAAT 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAAT 300

Qy        423 AGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTT 479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTT 357


RESULT 7
AC073930
LOCUS       AC073930               189729 bp    DNA     linear   HTG 01-MAR-2002
DEFINITION  Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC073930
VERSION     AC073930.23  GI:18958593
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 189729)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14861669.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBWS
                Center clone name: RP11-144B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191640 bases at least Q40
```

```
                Consensus quality: 195135 bases at least Q30
                Consensus quality: 197786 bases at least Q20
                Estimated insert size: 193340; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
             --------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 8 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1    64773: contig of 64773 bp in length
             *    64774    64873: gap of unknown length
             *    64874    99097: contig of 34224 bp in length
             *    99098    99197: gap of unknown length
             *    99198   135331: contig of 36134 bp in length
             *   135332   135431: gap of unknown length
             *   135432   161344: contig of 25913 bp in length
             *   161345   161444: gap of unknown length
             *   161445   171231: contig of 9787 bp in length
             *   171232   171331: gap of unknown length
             *   171332   180922: contig of 9591 bp in length
             *   180923   181022: gap of unknown length
             *   181023   187271: contig of 6249 bp in length
             *   187272   187371: gap of unknown length
             *   187372   189729: contig of 2358 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-144B2"
     gap             64774. .64873
                     /estimated_length=unknown
     gap             99098. .99197
                     /estimated_length=unknown
     gap             135332. .135431
                     /estimated_length=unknown
     gap             161345. .161444
                     /estimated_length=unknown
     gap             171232. .171331
                     /estimated_length=unknown
     gap             180923. .181022
                     /estimated_length=unknown
     gap             187272. .187371
                     /estimated_length=unknown
ORIGIN

  Query Match             44.6%;  Score 312;  DB 14;  Length 189729;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-74;
  Matches  312;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        390 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171332 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 171391

Qy        450 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171392 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 171451

Qy        510 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171452 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 171511

Qy        570 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171512 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 171571

Qy        630 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171572 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 171631

Qy        690 CAACCTCCTGAG 701
              ||||||||||||
Db     171632 CAACCTCCTGAG 171643


RESULT 8
AC158058/c
LOCUS       AC158058              227930 bp    DNA     linear   HTG 01-JUL-2005
DEFINITION  Bos taurus clone CH240-58P17, WORKING DRAFT SEQUENCE, 27 unordered
            pieces.
ACCESSION   AC158058
VERSION     AC158058.2  GI:68266594
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 227930)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 227930)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2005) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 227930)
  AUTHORS   Cow Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 28, 2005 this sequence version replaced gi:60498830.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: FDGU
                Center clone name: CH240-58P17
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 218051 bases at least Q40
                Consensus quality: 219911 bases at least Q30
                Consensus quality: 221398 bases at least Q20
                Estimated insert size: 220393; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6423: contig of 6423 bp in length
            *    6424     6574: gap of 151 bp
            *    6575    11190: contig of 4616 bp in length
            *   11191    11364: gap of 174 bp
            *   11365    32898: contig of 21534 bp in length
            *   32899    32998: gap of unknown length
            *   32999    57556: contig of 24558 bp in length
            *   57557    57606: gap of 50 bp
            *   57607    70177: contig of 12571 bp in length
            *   70178    70227: gap of 50 bp
            *   70228    80485: contig of 10258 bp in length
            *   80486    80535: gap of 50 bp
            *   80536    86613: contig of 6078 bp in length
            *   86614    86663: gap of 50 bp
            *   86664   105235: contig of 18572 bp in length
            *  105236   105298: gap of 63 bp
            *  105299   114241: contig of 8943 bp in length
            *  114242   114291: gap of 50 bp
            *  114292   115475: contig of 1184 bp in length
            *  115476   115525: gap of 50 bp
            *  115526   116869: contig of 1344 bp in length
            *  116870   117227: gap of 358 bp
            *  117228   123014: contig of 5787 bp in length
            *  123015   123171: gap of 157 bp
            *  123172   127191: contig of 4020 bp in length
            *  127192   127291: gap of unknown length
            *  127292   131796: contig of 4505 bp in length
            *  131797   131846: gap of 50 bp
            *  131847   143201: contig of 11355 bp in length
            *  143202   143251: gap of 50 bp
            *  143252   153087: contig of 9836 bp in length
            *  153088   153137: gap of 50 bp
            *  153138   162452: contig of 9315 bp in length
            *  162453   163639: gap of 1187 bp
            *  163640   178047: contig of 14408 bp in length
            *  178048   178097: gap of 50 bp
            *  178098   180637: contig of 2540 bp in length
            *  180638   181569: gap of 932 bp
            *  181570   195701: contig of 14132 bp in length
            *  195702   195751: gap of 50 bp
            *  195752   201668: contig of 5917 bp in length
            *  201669   201718: gap of 50 bp
            *  201719   213784: contig of 12066 bp in length
            *  213785   213834: gap of 50 bp
            *  213835   220883: contig of 7049 bp in length
            *  220884   220983: gap of unknown length
            *  220984   223706: contig of 2723 bp in length
            *  223707   224530: gap of 824 bp
            *  224531   225570: contig of 1040 bp in length
            *  225571   225670: gap of unknown length
            *  225671   226760: contig of 1090 bp in length
            *  226761   226860: gap of unknown length
            *  226861   227930: contig of 1070 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227930
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-58P17"
     gap             6424. .6574
                     /estimated_length=151
     gap             11191. .11364
                     /estimated_length=174
     gap             32899. .32998
                     /estimated_length=unknown
     gap             57557. .57606
                     /estimated_length=50
     gap             70178. .70227
                     /estimated_length=50
     gap             80486. .80535
                     /estimated_length=50
     gap             86614. .86663
                     /estimated_length=50
     gap             105236. .105298
                     /estimated_length=63
     gap             114242. .114291
                     /estimated_length=50
     gap             115476. .115525
                     /estimated_length=50
     gap             116870. .117227
                     /estimated_length=358
     gap             123015. .123171
                     /estimated_length=157
     gap             127192. .127291
                     /estimated_length=unknown
     gap             131797. .131846

  Query Match            30.2%;  Score 211.4;  DB 14;  Length 227930;
  Best Local Similarity  69.7%;  Pred. No. 1.4e-46;
  Matches  363;  Conservative    2;  Mismatches  133;  Indels   23;  Gaps    5;

Qy         93 TTTTTTCAAGCCATAGGATCATTTCTCCTGA-AGCATCTTGGCGAAGTCATCCCCACCTG 151
              | ||||||||||| |||| || | |  || | ||||  |   | ||| || |||||||||
```

```
Db      32892 TGTTTTCAAGCCACAGGACCAGTACCTCTCAGAGCAGGTCATCTAAGACAGCCCCACCTG 32833

Qy        152 TTCCTGAGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCT 211
              |||    || | ||| ||    ||||| | |  ||| || | | | |||||  |||||||
Db      32832 CTCC---AAGGGAGCAAGTTCAGGCTGGCGTGCTGCCCTACCCATGCTCCTGCCTCAGCT 32776

Qy        212 --GTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAG 269
                || ||||||  | ||||||||||||||||| || ||||||||||: | || ||||||
Db      32775 CCGTTCCTCCCCTTCTCCAGGTGCTGCCAGACGCACGACAACTGCTACAAACAAGCCAAG 32716

Qy        270 AAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCG 329
              || ||||||||||||||| | ||  |||||: || |||||| ||| ||| || |||||
Db      32715 AAACTGGACAGCTGTAAAGTCCTCGTGGACAATCCCTACACCAACAACTACTCCTACTCA 32656

Qy        330 TGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTC 389
              || |||  |   |  |||||||| |||||||||||| ||  |  |||
Db      32655 TGTTCTAACAATGAGATCACCTGCAGCAGTAGGTTTACCCTCTTGTTG------------ 32608

Qy        390 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 449
                | |  | |    |  || ||||||||||||||  | |||||| | | | || |
Db      32607 --GCTCTTGTTAGCGGGCATGGAGGGAAATAATAGTAATGATAGCCATCACTGACTGATC 32550

Qy        450 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 509
              |||||||||||  |  |||||| |||||||||||||||   ||||||  ||||| |||||||
Db      32549 ATTTTCTTGGTGCCAGGCAGTATCTCCTTTAATCCTTGAAACAACTGTATGGAATAGGTA 32490

Qy        510 C---AATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGC 566
              |    ||||||| || || |||||| ||  |||||||||||||||| |||  | | |||
Db      32489 CTGTTATTATCCCCATTTTACAGATGAGGCAACTGAGGCTCAGAAGCCTGGACAAATTGT 32430

Qy        567 CCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTT 607
              |||||  |||||||       |    |      |||||| ||||
Db      32429 CCAAGGGCACACAGTCCCATAAATGGTGGTGTTTGGTTTTT 32389
```

RESULT 9
AC127942/c
LOCUS AC127942 215223 bp DNA linear HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-142L1, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION AC127942
VERSION AC127942.2 GI:23907779
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 215223)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215223)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 215223)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 12, 2002 this sequence version replaced gi:21908474.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GXSJ
Center clone name: CH230-142L1
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 200801 bases at least Q40
Consensus quality: 203059 bases at least Q30
Consensus quality: 204719 bases at least Q20
Estimated insert size: 206878; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    28323: contig of 28323 bp in length
          *    28324    28423: gap of unknown length
          *    28424    55628: contig of 27205 bp in length
          *    55629    55728: gap of unknown length
          *    55729   213935: contig of 158207 bp in length
          *   213936   214035: gap of unknown length
          *   214036   215223: contig of 1188 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215223
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-142L1"
     misc_feature    complement(6631. .6964)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWAXO61TJB"
     gap             28324. .28423
                     /estimated_length=unknown
     misc_feature    28424. .29841
                     /note="wgs_contig"
     gap             55629. .55728
                     /estimated_length=unknown
     misc_feature    55729. .58010
                     /note="wgs_contig"
     gap             213936. .214035
                     /estimated_length=unknown
ORIGIN

  Query Match             20.1%;  Score 140.6;  DB 14;  Length 215223;
  Best Local Similarity   65.0%;  Pred. No. 2.8e-27;
  Matches  236;  Conservative    3;  Mismatches  117;  Indels    7;  Gaps    2;

Qy         53 AAAAAATGCTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATC 112
              ||||  ||||||||||||||| |||| | | ||   | |||  |  || ||||   ||  |
Db      25680 AAAAGTTGCTTTCAATAAATCTATGCTGATAGAGATCATAGATCCTCGAGCCTCGGGGCC 25621

Qy        113 ATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGA 172
              |||    ||| ||||   ||   |||| | || |   ||||    | ||   | ||||||
Db      25620 ATTAGCTCTGGAGCACTGTGTTTAAGTTAGCCTCGAATGTTTGAAAAAGCCCGTAGGTGA 25561

Qy        173 GGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTG 232
              |  | ||| |   |   | |||      |      |  | |  | |||  ||  |||||||
Db      25560 GAACCGACTTCCGGTGGTCCAC-----GCCTCTCCTTCCGCTCTTCCTTCTCGCCAGGTG 25506

Qy        233 CTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCT 292
              ||||||||| |||||| |||||||: | |||||||||||||||| |||||||||| ||
Db      25505 CTGCCAGACTCATGACCACTGCTACAATCAGGCCAAGAAGCTGGAAAGCTGTAAATTCCT 25446

Qy        293 GCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTG 352
                | ||||::|| |||||| |||| || |||||| |||||| ||    | ||||||||
Db      25445 CATCGACAACCCCTACACCAACACGTACTCATACAAGTGCTCCGGGAACGTGATCACCTG 25386

Qy        353 TAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGG 412
              ||| ||||| ||| || ||| ||||| || | |||   |||  |  |  || | ||
Db      25385 CAGCGGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCCTGGCCTTTGGCCGGTCTGG 25328

Qy        413 GGG 415
                ||
Db      25327 AGG 25325


RESULT 10
AC097575
LOCUS       AC097575               238961 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-20A7, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC097575
VERSION     AC097575.6  GI:30581004
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 238961)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238961)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238961)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23096385.
```

```
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GIUS
              Center clone name: CH230-20A7
          -------------- Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 222002 bases at least Q40
              Consensus quality: 225364 bases at least Q30
              Consensus quality: 227894 bases at least Q20
              Estimated insert size: 232967; sum-of-contigs estimation
              Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   234563: contig of 234563 bp in length
          *   234564   234663: gap of unknown length
          *   234664   236006: contig of 1343 bp in length
          *   236007   236106: gap of unknown length
          *   236107   237379: contig of 1273 bp in length
          *   237380   237479: gap of unknown length
          *   237480   238961: contig of 1482 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238961
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-20A7"
     misc_feature    1. .1200
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    15551. .15691
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH317687"
     misc_feature    21638. .67975
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH317686"
     misc_feature    79112. .81297
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    233264. .234563
                     /note="wgs_end_extension
                     clone_end:T7"
     gap             234564. .234663
                     /estimated_length=unknown
     gap             236007. .236106
                     /estimated_length=unknown
     gap             237380. .237479
                     /estimated_length=unknown
ORIGIN

  Query Match             20.1%;  Score 140.6;  DB 14;  Length 238961;
  Best Local Similarity   65.0%;  Pred. No. 2.8e-27;
  Matches  236;  Conservative    3;  Mismatches  117;  Indels    7;  Gaps    2;

Qy         53 AAAAAATGCTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATC 112
              ||||  |||||||||||||||| |||| | | ||   | |||  |  || ||||   ||  |
Db      73855 AAAAGTTGCTTTCAATAAATCTATGCTGATAGAGATCATAGATCCTCGAGCCTCGGGGCC 73914

Qy        113 ATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGA 172
              |||    ||| ||||   ||   |||| | || |   ||||    | ||   | ||||||
Db      73915 ATTAGCTCTGGAGCACTGTGTTTAAGTTAGCCTCGAATGTTTGAAAAAGCCCGTAGGTGA 73974

Qy        173 GGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTG 232
              | | ||| |   |  | |||      |    | | |  | |||  ||  |||||||
Db      73975 GAACCGACTTCCGGTGGTCCAC-----GCCTCTCCTTCCGCTCTTCCTTCTCGCCAGGTG 74029

Qy        233 CTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCT 292
              |||||||||  |||||| |||||||: | ||||||||||||||||| ||||||||||| ||
Db      74030 CTGCCAGACTCATGACCACTGCTACAATCAGGCCAAGAAGCTGGAAAGCTGTAAATTCCT 74089

Qy        293 GCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTG 352
                | ||||::|| |||||| |||| || |||||| |||||| ||    |  ||||||||
Db      74090 CATCGACAACCCCTACACCAACACGTACTCATACAAGTGCTCCGGGAACGTGATCACCTG 74149

Qy        353 TAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGG 412
              ||| ||||| |||  || |||  ||||| || | |||   |||   |  |  || | ||
Db      74150 CAGCGGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCCTGGCCTTTGGCCGGTCTGG 74207

Qy        413 GGG 415
              |||
Db      74208 AGG 74210


RESULT 11
AC096966/c
LOCUS       AC096966              290053 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-213K13, *** SEQUENCING IN PROGRESS
            ***, 12 unordered pieces.
ACCESSION   AC096966
VERSION     AC096966.6  GI:30521249
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 290053)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
```

```
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 290053)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 290053)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23268834.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIFJ
                Center clone name: CH230-213K13
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 249124 bases at least Q40
                Consensus quality: 252995 bases at least Q30
                Consensus quality: 256136 bases at least Q20
                Estimated insert size: 261662; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27647: contig of 27647 bp in length
            *    27648    27747: gap of unknown length
            *    27748    44239: contig of 16492 bp in length
            *    44240    44339: gap of unknown length
            *    44340   271309: contig of 226970 bp in length
            *   271310   271409: gap of unknown length
            *   271410   272830: contig of 1421 bp in length
            *   272831   272930: gap of unknown length
            *   272931   274121: contig of 1191 bp in length
            *   274122   274221: gap of unknown length
            *   274222   275624: contig of 1403 bp in length
            *   275625   275724: gap of unknown length
            *   275725   277023: contig of 1299 bp in length
            *   277024   277123: gap of unknown length
            *   277124   278863: contig of 1740 bp in length
            *   278864   278963: gap of unknown length
            *   278964   280726: contig of 1763 bp in length
            *   280727   280826: gap of unknown length
            *   280827   284233: contig of 3407 bp in length
            *   284234   284333: gap of unknown length
            *   284334   286731: contig of 2398 bp in length
            *   286732   286831: gap of unknown length
            *   286832   290053: contig of 3222 bp in length.
FEATURES             Location/Qualifiers
     source          1. .290053
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-213K13"
     gap             27648. .27747
                     /estimated_length=unknown
     misc_feature    35097. .36658
                     /note="wgs_contig"
     gap             44240. .44339
                     /estimated_length=unknown
     misc_feature    44340. .45348
                     /note="wgs_contig"
     misc_feature    270030. .271309
                     /note="wgs_contig"
     gap             271310. .271409
                     /estimated_length=unknown
     gap             272831. .272930
                     /estimated_length=unknown
     gap             274122. .274221
                     /estimated_length=unknown
     gap             275625. .275724
                     /estimated_length=unknown
     gap             277024. .277123
                     /estimated_length=unknown
     gap             278864. .278963
                     /estimated_length=unknown
     gap             280727. .280826
                     /estimated_length=unknown
     gap             284234. .284333
                     /estimated_length=unknown
     gap             286732. .286831
                     /estimated_length=unknown
ORIGIN

  Query Match             20.1%;  Score 140.6;  DB 14;  Length 290053;
  Best Local Similarity   65.0%;  Pred. No. 2.8e-27;
  Matches  236;  Conservative    3;  Mismatches  117;  Indels    7;  Gaps    2;

Qy         53 AAAAAATGCTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATC 112
```

```
Db      65775 AAAAGTTGCTTTCAATAAATCTATGCTGATAGAGATCATAGATCCTCGAGCCTCGGGGCC 65716

Qy        113 ATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGA 172

Db      65715 ATTAGCTCTGGAGCACTGTGTTTAAGTTAGCCTCGAATGTTTGAAAAAGCCCGTAGGTGA 65656

Qy        173 GGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTG 232

Db      65655 GAACCGACTTCCGGTGGTCCAC-----GCCTCTCCTTCCGCTCTTCCTTCTCGCCAGGTG 65601

Qy        233 CTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCT 292

Db      65600 CTGCCAGACTCATGACCACTGCTACAATCAGGCCAAGAAGCTGGAAAGCTGTAAATTCCT 65541

Qy        293 GCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTG 352

Db      65540 CATCGACAACCCCTACACCAACACGTACTCATACAAGTGCTCCGGGAACGTGATCACCTG 65481

Qy        353 TAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGG 412

Db      65480 CAGCGGTAGGGTTA--CCCTCCGGGACCTCTGGACTCTGCCTGGCCTTTGGCCGGTCTGG 65423

Qy        413 GGG 415

Db      65422 AGG 65420


RESULT 12
CQ922146
LOCUS       CQ922146                418 bp    DNA     linear   PAT 23-NOV-2004
DEFINITION  Sequence 3346 from Patent WO2004097052.
ACCESSION   CQ922146
VERSION     CQ922146.1  GI:56212087
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burczynski,M.E., Twine,N.C., Slonim,D.K., Trepicchio,W.L.,
            Strahs,A., Immerman,F. and Dorner,A.J.
  TITLE     Methods for prognosis and treatment of solid tumors
  JOURNAL   Patent: WO 2004097052-A 3346 11-NOV-2004;
            Wyeth (US); Burczynski, Michael E. (US)
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 6;  Length 418;
  Best Local Similarity   94.3%;  Pred. No. 1.5e-24;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280

Db        185 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAG 244

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340

Db        245 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 304

Qy        341 GGCAATCACCTGTAGCAGTA 360

Db        305 GGCAATCACCTGTAGCAGCA 324


RESULT 13
AX377240
LOCUS       AX377240                447 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 2 from Patent WO0212562.
ACCESSION   AX377240
VERSION     AX377240.1  GI:19573529
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 2 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 6;  Length 447;
  Best Local Similarity   94.3%;  Pred. No. 1.5e-24;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280

Db        186 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 245

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340

Db        246 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 305

Qy        341 GGCAATCACCTGTAGCAGTA 360

Db        306 GGCAATCACCTGTAGCAGCA 325


RESULT 14
CQ723247
LOCUS       CQ723247                559 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 9181 from Patent WO02068579.
ACCESSION   CQ723247
VERSION     CQ723247.1  GI:42284104
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 9181 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 6;  Length 559;
  Best Local Similarity   94.3%;  Pred. No. 1.5e-24;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280

Db        221 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 280

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
```

```
                    ||||||||||||||||||||::||||||||||||||||||||||||||||||||||||||||
Db        281 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340

Qy        341 GGCAATCACCTGTAGCAGTA 360
                  |||||||||||||||||| |
Db        341 GGCAATCACCTGTAGCAGCA 360


RESULT 15
AR380655
LOCUS       AR380655                562 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 1200 from patent US 6607879.
ACCESSION   AR380655
VERSION     AR380655.1  GI:40088289
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Cocks,B.G., Stuart,S.G. and Seilhamer,J.J.
  TITLE     Compositions for the detection of blood cell and immunological
            response gene expression
  JOURNAL   Patent: US 6607879-A 1200 19-AUG-2003;
            Incyte Corporation; Palo Alto, CA
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 6;  Length 562;
  Best Local Similarity   94.3%;  Pred. No. 1.5e-24;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
                  |||   || ||||||||||||||||||||||||||:||||||||||||||||||||
Db        223 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 282

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
                  ||||||||||||||||||::||||||||||||||||||||||||||||||||||||||
Db        283 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 342

Qy        341 GGCAATCACCTGTAGCAGTA 360
                  |||||||||||||||||| |
Db        343 GGCAATCACCTGTAGCAGCA 362


Search completed: February  9, 2006, 20:42:23
Job time : 3498.02 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 451.558 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 699.8
Sequence: 1 gtcgtgtcactgctgtccag.....tgcagcctcaacctcctgag 701

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	699.8	100.0	12174	12	ADJ09983	Adj09983 Human pho
2	699.8	100.0	13612	6	ABK47376	Abk47376 Human Pho
3	699.4	100.0	12174	12	ADI35082	Adi35082 Human PLA
4	380.8	54.4	3375	3	AAA35101	Aaa35101 Human ade
5	380.8	54.4	3375	3	AAF21223	Aaf21223 Human low
6	380.8	54.4	3375	10	ABZ96917	Abz96917 Human nuc
7	380.8	54.4	3375	11	ABD20766	Abd20766 Human pul
8	380.8	54.4	6771	3	AAA35103	Aaa35103 Human ade
9	380.8	54.4	6771	3	AAF21225	Aaf21225 Human low
10	380.8	54.4	6771	10	ABZ96919	Abz96919 Human nuc
11	380.8	54.4	6771	11	ABD20768	Abd20768 Human pul
12	130.8	18.7	418	13	ADU12907	Adu12907 Solid tum
13	130.8	18.7	447	6	ABK47377	Abk47377 Human Pho
14	130.8	18.7	447	14	ADV43454	Adv43454 Human psy
15	130.8	18.7	562	3	AAA35100	Aaa35100 Human ade
16	130.8	18.7	562	3	AAF21222	Aaf21222 Human low
17	130.8	18.7	562	6	ABL64906	Abl64906 Lung canc
18	130.8	18.7	562	6	ABL65512	Abl65512 Lung canc
19	130.8	18.7	562	6	ABL65915	Abl65915 Lung canc
20	130.8	18.7	562	10	ABZ96916	Abz96916 Human nuc
21	130.8	18.7	562	11	ADI31874	Adi31874 Human cDN
22	130.8	18.7	562	11	ABD20765	Abd20765 Human pul
23	130.8	18.7	562	13	ADS83941	Ads83941 Human lym
24	130.8	18.7	585	12	ADO28598	Ado28598 Human PA2
25	130.8	18.7	630	3	AAC98824	Aac98824 Human pan
26	130.8	18.7	644	2	AAZ41391	Aaz41391 Human nor
27	130.8	18.7	654	4	AAH57532	Aah57532 Human pan
28	130.8	18.7	654	5	AAS76099	Aas76099 DNA encod
29	130.8	18.7	742	13	ACN42348	Acn42348 Human dia
c 30	126.2	18.0	496	4	AAH57324	Aah57324 Human pan
c 31	126	18.0	26593	13	ABD33464	Abd33464 Murine ca
32	121.8	17.4	445	3	AAC99186	Aac99186 Human pan
33	89.8	12.8	369	1	AAN81342	Aan81342 DNA encod
c 34	86.6	12.4	223556	11	ACN44110	Acn44110 Human gen
35	86.6	12.4	243428	12	ADF51132	Adf51132 Human P-R
36	85.6	12.2	96596	9	ADA02834	Ada02834 Human MEF
37	85.6	12.2	96596	10	ADB72572	Adb72572 Human MEF
38	85.6	12.2	96596	10	ADC85313	Adc85313 Mouse Mef
39	85.6	12.2	96596	12	ADM74429	Adm74429 Human car
40	85	12.1	460	2	AAQ04439	Aaq04439 Gene enco
41	83.4	11.9	460	1	AAN91424	Aan91424 DNA encod
42	81.6	11.7	34551	10	ADC87424	Adc87424 Human GPC
43	81	11.6	3509	11	ADM02028	Adm02028 Human cDN
44	80	11.4	62658	13	ABD33339	Abd33339 Human can
45	80	11.4	143306	6	ABK49586	Abk49586 Human tra

ALIGNMENTS

RESULT 1
ADJ09983
ID ADJ09983 standard; DNA; 12174 BP.
XX
AC ADJ09983;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(436,c)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(839,a)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4050,a)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4689,t)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6282,a)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6358,c)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6653,t)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7256,t)
FT /*tag= h

```
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7300,a)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7301,c)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7328,g)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(8062,c)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(9182,t)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(11649,c)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphisms"
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   P-PSDB; ADJ09984.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ   Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 699.8;  DB 12;  Length 12174;
  Best Local Similarity   99.6%;  Pred. No. 4.4e-191;
  Matches  698;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7000 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 7059

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7060 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 7119

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7120 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 7179

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7180 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db       7240 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 7299

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7300 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7360 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 7419

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7420 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 7479

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7480 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7539

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7540 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7599

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7600 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7659

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db       7660 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 7700


RESULT 2
ABK47376
ID   ABK47376 standard; DNA; 13612 BP.
XX
AC   ABK47376;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW   Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(3845,A)
FT                   /*tag=  a
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(3968,A)
FT                   /*tag=  b
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4053. .9613
FT                   /*tag=  c
FT                   /product= "Pla2G1B"
FT   exon            4053. .4086
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4087. .5785
```

```
FT                   /*tag=  e
FT                   /number= 1
FT   exon            5786. .5945
FT                   /*tag=  f
FT                   /number= 2
FT   intron          5946. .6744
FT                   /*tag=  g
FT                   /number= 2
FT   variation       replace(6060,A)
FT                   /*tag=  h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   exon            6745. .6873
FT                   /*tag=  i
FT                   /number= 3
FT   variation       replace(6844,A)
FT                   /*tag=  j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   intron          6874. .9489
FT                   /*tag=  k
FT                   /number= 3
FT   exon            9490. .9613
FT                   /*tag=  l
FT                   /number= 4
FT   variation       replace(9531,A)
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 699.8;  DB 6;  Length 13612;
  Best Local Similarity   99.6%;  Pred. No. 4.7e-191;
  Matches  698;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6816 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7116 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db       7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 7216


RESULT 3
ADI35082
ID   ADI35082 standard; DNA; 12174 BP.
XX
AC   ADI35082;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human PLA2G1B nucleotide sequence.
XX
```

KW PLA2G1B ; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic
CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC subject, where the presence of the polymorphic variation indicates a
CC predisposition to fat deposition in the subject. The polymorphic
CC variation is a guanine at position 7328 or thymine at position 9182 of
CC the present sequence. The method is useful for diagnosing a
CC predisposition to fat deposition or leanness in a subject, and
CC consequently for diagnosing a predisposition to non-insulin dependent
CC diabetes mellitus (NIDDM) in a subject and conditions such as
CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC hypercoagulability, or microalbuminuria, which can lead to early
CC prescription of preventive measures. The present sequence represents a
CC human PLA2G1B nucleotide sequence.
XX
SQ Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

Query Match 100.0%; Score 699.8; DB 12; Length 12174;
Best Local Similarity 99.9%; Pred. No. 5.8e-191;
Matches 700; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7000 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 7059

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7060 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 7119

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7120 TGAAGCATCTTGGCGAAGTCATCCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 7179

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7180 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7240 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 7299

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db       7300 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7360 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 7419

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7420 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 7479

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7480 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7539

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7540 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7599

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7600 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7659

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||||||||||||||||||||||||||||||||||||||||
Db       7660 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 7700
```

RESULT 4
AAA35101
ID AAA35101 standard; DNA; 3375 BP.
XX
AC AAA35101;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2790.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 1056-1057; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,

CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing
XX
SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 54.4%; Score 380.8; DB 3; Length 3375;
Best Local Similarity 99.2%; Pred. No. 4.1e-99;
Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3053

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3113

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3173

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3233

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3234 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3293

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3353

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||||
Db       3354 GGTTTATCCCTTCCTTGACCTA 3375
```

RESULT 5
AAF21223
ID AAF21223 standard; DNA; 3375 BP.
XX
AC AAF21223;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2790.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX
PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
XX
PS Disclosure; Page 1135-1136; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 54.4%; Score 380.8; DB 3; Length 3375;
Best Local Similarity 99.2%; Pred. No. 4.1e-99;
Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3053

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3113

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3173

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3233

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
                |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3293

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
                ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3353

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
                |||||||||||||||||||||||
Db       3354 GGTTTATCCCTTCCTTGACCTA 3375
```

```
RESULT 6
ABZ96917
ID   ABZ96917 standard; DNA; 3375 BP.
XX
AC   ABZ96917;
XX
DT   17-OCT-2003  (first entry)
XX
DE   Human nucleic acid sequence.
XX
KW   Human; antisense; lung dysfunction; nasal airway dysfunction;
KW   antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW   antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW   antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW   adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW   lung inflammation; respiratory disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285308-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013135.
XX
PR   24-APR-2001; 2001US-0286137P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-229219/22.
XX
PT   Pharmaceutical composition for treating ailments associated with impaired
PT   respiration, has oligo(s) antisense to specific gene(s) or its
PT   corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT   ubiquinone.
XX
PS   Disclosure; SEQ ID NO 12159; 872pp; English.
XX
CC   The invention relates to a novel pharmaceutical composition, which has a
CC   first active agent comprising an oligonucleotide antisense to the
CC   initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC   5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC   junctions of genes encoding a polypeptide associated with lung and/or
CC   nasal airway dysfunction and a second active agent comprising an
CC   antiinflammatory steroid and ubiquinone. A composition of the invention
CC   has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC   immunosuppressive, and cytostatic activity. The composition may have a
CC   use in antisense gene therapy. The composition is useful for treating or
CC   preventing a respiratory, lung or malignant disease or condition, also
CC   for enhancing the prophylactic or therapeutic respiratory effect of an
CC   antiinflammatory steroid in a subject, for reducing or depleting levels
CC   of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC   receptor, producing bronchodilation, increasing levels of ubiquinone or
CC   lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC   lung inflammation, lung allergies, or a respiratory disease or condition.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

  Query Match             54.4%;  Score 380.8;  DB 10;  Length 3375;
  Best Local Similarity   99.2%;  Pred. No. 4.1e-99;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3053

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3113

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3173

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3233

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3293

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3353

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||||
Db       3354 GGTTTATCCCTTCCTTGACCTA 3375
```

```
RESULT 7
ABD20766
ID   ABD20766 standard; DNA; 3375 BP.
XX
AC   ABD20766;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human pulmonary and inflammatory target DNA #377.
XX
KW   Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW   respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW   surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW   analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW   beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW   respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW   emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW   pulmonary transplantation rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285309-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013143.
XX
PR   24-APR-2001; 2001US-0286036P.
XX
```

PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-093058/08.
XX
PT Pharmaceutical composition for treating asthma, has antisense
PT oligonucleotide containing less percentage of adenosine, targeted to
PT nucleic acids associated with lung airway or lung dysfunction, and
PT bronchodilating agent.
XX
PS Claim 15; SEQ ID NO 12159; 763pp; English.
XX
CC This invention describes a novel composition (a) a first active agent,
CC comprising oligonucleotides, effective for alleviating
CC bronchoconstriction, respiratory tract inflammation, allergies and
CC reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC surfactant depletion or hyposecretion, when administered to a mammal. The
CC oligonucleotides are derived from a gene encoding or regulating
CC expression of a target polypeptide associated with lung airway or lung
CC dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC The invention also describes a kit, that comprises: (a) a delivery
CC device, in separate containers, (b) the oligonucleotides, (c)
CC instructions for adding a carrier and for use of the kit. The composition
CC of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC beta-adrenergic agonist. The composition is useful for preventing or
CC treating a respiratory, lung or malignant disease. The administered
CC composition comprises oligo and is administered to reduce the production
CC or availability, or to increase the degradation of the target mRNA or to
CC reduce the amount of target polypeptide present in the lungs. The
CC pulmonary obstruction, and/or bronchoconstriction and/or lung
CC inflammation, allergies and/or surfactant hypoproduction are associated
CC with a disease or condition such as pulmonary vasoconstriction,
CC inflammation, allergies, asthma, impeded respiration, respiratory
CC distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC transplantation rejection, pulmonary infections, bronchitis or cancer.
CC The reduced adenosine content of the anti-sense oligos corresponding to
CC thymidines present in the target RNA serves to prevent the breakdown of
CC the oligonucleotides into products that free adenosine into the system
CC e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC prevent any unwanted effects due to it
XX
SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 54.4%; Score 380.8; DB 11; Length 3375;
Best Local Similarity 99.2%; Pred. No. 4.1e-99;
Matches 379; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3053

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3113

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3173

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3233

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3234 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3293

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3353

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              ||||||||||||||||||||||
Db       3354 GGTTTATCCCTTCCTTGACCTA 3375
```

RESULT 8
AAA35103
ID AAA35103 standard; DNA; 6771 BP.
XX
AC AAA35103;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2792.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 1058-1059; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence

```
CC   listing
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             54.4%;  Score 380.8;  DB 3;  Length 6771;
  Best Local Similarity   99.2%;  Pred. No. 5.6e-99;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3556 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 3615

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3616 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3675

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3676 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3735

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3736 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3795

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3796 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3855

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3856 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3915

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||||
Db       3916 GGTTTATCCCTTCCTTGACCTA 3937


RESULT 9
AAF21225
ID   AAF21225 standard; DNA; 6771 BP.
XX
AC   AAF21225;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2792.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US008020.
XX
PR   06-APR-1999;   99US-0127958P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not trigger
PT   adenosine receptors during metabolism, useful e.g. for treating cancers
PT   and respiratory obstructions.
XX
PS   Disclosure; Page 1137-1139; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC   surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             54.4%;  Score 380.8;  DB 3;  Length 6771;
  Best Local Similarity   99.2%;  Pred. No. 5.6e-99;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3556 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 3615

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3616 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3675

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3676 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3735

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3736 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3795

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3796 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3855

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3856 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3915

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||||
Db       3916 GGTTTATCCCTTCCTTGACCTA 3937
```

```
RESULT 10
ABZ96919
ID   ABZ96919 standard; DNA; 6771 BP.
XX
AC   ABZ96919;
XX
DT   17-OCT-2003  (first entry)
XX
DE   Human nucleic acid sequence.
XX
KW   Human; antisense; lung dysfunction; nasal airway dysfunction;
KW   antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW   antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW   antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW   adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW   lung inflammation; respiratory disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285308-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013135.
XX
PR   24-APR-2001; 2001US-0286137P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-229219/22.
XX
PT   Pharmaceutical composition for treating ailments associated with impaired
PT   respiration, has oligo(s) antisense to specific gene(s) or its
PT   corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT   ubiquinone.
XX
PS   Disclosure; SEQ ID NO 12161; 872pp; English.
XX
CC   The invention relates to a novel pharmaceutical composition, which has a
CC   first active agent comprising an oligonucleotide antisense to the
CC   initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC   5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC   junctions of genes encoding a polypeptide associated with lung and/or
CC   nasal airway dysfunction and a second active agent comprising an
CC   antiinflammatory steroid and ubiquinone. A composition of the invention
CC   has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC   immunosuppressive, and cytostatic activity. The composition may have a
CC   use in antisense gene therapy. The composition is useful for treating or
CC   preventing a respiratory, lung or malignant disease or condition, also
CC   for enhancing the prophylactic or therapeutic respiratory effect of an
CC   antiinflammatory steroid in a subject, for reducing or depleting levels
CC   of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC   receptor, producing bronchodilation, increasing levels of ubiquinone or
CC   lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC   lung inflammation, lung allergies, or a respiratory disease or condition.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             54.4%;  Score 380.8;  DB 10;  Length 6771;
  Best Local Similarity   99.2%;  Pred. No. 5.6e-99;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3556 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3615

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3616 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3675

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3676 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3735

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3736 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3795

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       3796 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3855

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3856 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3915

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||
Db       3916 GGTTTATCCCTTCCTTGACCTA 3937


RESULT 11
ABD20768
ID   ABD20768 standard; DNA; 6771 BP.
XX
AC   ABD20768;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human pulmonary and inflammatory target DNA #379.
XX
KW   Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW   respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW   surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW   analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW   beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW   respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW   emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW   pulmonary transplantation rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285309-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013143.
XX
PR   24-APR-2001; 2001US-0286036P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-093058/08.
XX
PT   Pharmaceutical composition for treating asthma, has antisense
PT   oligonucleotide containing less percentage of adenosine, targeted to
PT   nucleic acids associated with lung airway or lung dysfunction, and
PT   bronchodilating agent.
XX
PS   Claim 15; SEQ ID NO 12161; 763pp; English.
XX
CC   This invention describes a novel composition (a) a first active agent,
CC   comprising oligonucleotides, effective for alleviating
CC   bronchoconstriction, respiratory tract inflammation, allergies and
CC   reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC   surfactant depletion or hyposecretion, when administered to a mammal. The
CC   oligonucleotides are derived from a gene encoding or regulating
```

```
CC   expression of a target polypeptide associated with lung airway or lung
CC   dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC   The invention also describes a kit, that comprises: (a) a delivery
CC   device, in separate containers, (b) the oligonucleotides, (c)
CC   instructions for adding a carrier and for use of the kit. The composition
CC   of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC   analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC   beta-adrenergic agonist. The composition is useful for preventing or
CC   treating a respiratory, lung or malignant disease. The administered
CC   composition comprises oligo and is administered to reduce the production
CC   or availability, or to increase the degradation of the target mRNA or to
CC   reduce the amount of target polypeptide present in the lungs. The
CC   pulmonary obstruction, and/or bronchoconstriction and/or lung
CC   inflammation, allergies and/or surfactant hypoproduction are associated
CC   with a disease or condition such as pulmonary vasoconstriction,
CC   inflammation, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC   transplantation rejection, pulmonary infections, bronchitis or cancer.
CC   The reduced adenosine content of the anti-sense oligos corresponding to
CC   thymidines present in the target RNA serves to prevent the breakdown of
CC   the oligonucleotides into products that free adenosine into the system
CC   e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC   prevent any unwanted effects due to it
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             54.4%;  Score 380.8;  DB 11;  Length 6771;
  Best Local Similarity   99.2%;  Pred. No. 5.6e-99;
  Matches  379;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3556 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3615

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3616 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3675

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3676 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3735

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3736 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3795

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3796 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3855

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3856 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3915

Qy        361 GGTTTATCCCTTCCTTGACCTA 382
              |||||||||||||||||||||
Db       3916 GGTTTATCCCTTCCTTGACCTA 3937


RESULT 12
ADU12907
ID   ADU12907 standard; DNA; 418 BP.
XX
AC   ADU12907;
XX
DT   27-JAN-2005  (first entry)
XX
DE   Solid tumour prognosis gene seqid 3346.
XX
KW   cytostatic; gene therapy; expression profile; solid tumour;
KW   peripheral blood mononuclear cell; PBMC; prognosis; ds.
XX
OS   Unidentified.
XX
PN   WO2004097052-A2.
XX
PD   11-NOV-2004.
XX
PF   29-APR-2004; 2004WO-US013587.
XX
PR   29-APR-2003; 2003US-0466067P.
PR   23-JAN-2004; 2004US-0538246P.
XX
PA   (AMHP ) WYETH.
PA   (STRA/) STRAHS A.
XX
PI   Strahs A,  Trepicchio WL,  Burczynski ME,  Twine NC,  Slonim DK;
PI   Immerman F,  Dorner AJ;
XX
DR   WPI; 2004-804779/79.
XX
PT   A method, useful for prognosing and treating solid tumor, comprises
PT   comparing an expression profile of a gene expressed in peripheral blood
PT   mononuclear cells to a reference expression profile of a gene.
XX
PS   Disclosure; Page; 111pp; English.
XX
CC   The invention describes a method comprising comparing an expression
CC   profile of at least one gene in a peripheral blood sample of a patient to
CC   at least one reference expression profile of the at least one gene, where
CC   the patient has a solid tumour, and each of the gene is differentially
CC   expressed in peripheral blood mononuclear cells (PBMCs) of a first class
CC   of patients as compared to PBMCs of a second class of patients, where
CC   both the first and second classes of patients have the solid tumour, and
CC   each of the first and second classes is a subcluster formed by an
CC   unsupervised clustering analysis of gene expression profiles in PBMCs of
CC   a population of patients who have the solid tumour, and where the
CC   majority of the first class of patients has a first clinical outcome, and
CC   the majority of the second class of patients has a second clinical
CC   outcome. Also described are: a system comprising (i) a memory or a
CC   storage medium including data that represent an expression profile of at
CC   least one gene in a peripheral blood sample of a patient who has a solid
CC   tumour, (ii) at least another storage medium including data that
CC   represent at least one reference expression profile of the gene, (iii) a
CC   program capable of comparing the expression profile to the reference
CC   expression profile, and (iv) a processor capable of executing the
CC   program, where expression levels of the gene in peripheral blood
CC   mononuclear cells of patients who have the solid tumour correlate with
CC   clinical outcomes of the patients; and a nucleic acid or protein array
CC   comprising concentrated probes for solid tumour prognosis genes, where
CC   each of the solid tumour prognosis genes is differentially expressed in
CC   PBMCs of a first class of patients as compared to PBMCs of a second class
CC   of patients, where both the first and second classes of patients have a
CC   solid tumour, and where the first class of patients has a first clinical
CC   outcome, and the second class of patients has a second clinical outcome.
CC   The method, system, and array are useful for prognosing and treating
CC   solid tumours. This sequence represents a solid tumour prognosis gene of
CC   the invention. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 418 BP; 105 A; 125 C; 104 G; 84 T; 0 U; 0 Other;

  Query Match             18.7%;  Score 130.8;  DB 13;  Length 418;
  Best Local Similarity   94.3%;  Pred. No. 2.9e-27;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || ||||||||||||||||||||||||||:||||||||||||||||||||||
Db        185 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAG 244

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        245 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 304
```

```
Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        305 GGCAATCACCTGTAGCAGCA 324


RESULT 13
ABK47377
ID   ABK47377 standard; cDNA; 447 BP.
XX
AC   ABK47377;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B) cDNA.
XX
KW   Human; ss; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .447
FT                   /*tag= a
FT                   /product= "Pla2G1B"
FT   variation       replace(294,A)
FT                   /*tag= h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(365,A)
FT                   /*tag= j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 27; Fig 2; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The PLA2G1B gene is located on chromosome 12q23-q24.1.
CC   The present sequence is the cDNA encoding PLA2G1B
XX
SQ   Sequence 447 BP; 116 A; 128 C; 112 G; 91 T; 0 U; 0 Other;

  Query Match             18.7%;  Score 130.8;  DB 6;  Length 447;
  Best Local Similarity   94.3%;  Pred. No. 3e-27;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||:|||||||||||||||||||||
Db        186 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 245

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||
Db        246 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 305

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        306 GGCAATCACCTGTAGCAGCA 325


RESULT 14
ADV43454
ID   ADV43454 standard; cDNA; 447 BP.
XX
AC   ADV43454;
XX
DT   10-MAR-2005  (first entry)
XX
DE   Human psychoneuroendocrinimmune expressed sequence tag SEQ ID NO 1082.
XX
KW   microarray; psychoneuroendocrinimmune; chronic fatigue;
KW   non-insulin dependent diabetes; allergy; immune disorder; inflammation;
KW   cancer; neoplasm; infection; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004108899-A2.
XX
PD   16-DEC-2004.
XX
PF   04-JUN-2004; 2004WO-US017686.
XX
PR   04-JUN-2003; 2003US-0475915P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Nicholson A,  Vernon SD;
XX
DR   WPI; 2005-031682/03.
XX
PT   New microarray comprising probes for genes involved in
PT   psychoneuroendocrinimmune (PNI) activity, useful in diagnosing a
PT   condition associated with PNI activity, e.g., inflammatory or infectious
PT   diseases.
XX
PS   Claim 1; SEQ ID NO 1082; 254pp; English.
XX
CC   The invention relates to a new microarray which comprises probes for
CC   genes involved in psychoneuroendocrinimmune (PNI) activity. The
CC   microarray is useful in diagnosing a condition associated with PNI
CC   activity, such as CFS, type-2 diabetes, allergic condition, inflammation,
CC   cancer and infection. The present sequence represents a
```

```
CC   psychoneuroendocrinimmune gene expressed sequence tag. Note the
CC   specificatio mentions SEQ ID NO of up to 3314 but only sequences up to
CC   SEQ ID NO 1829 are provided.
XX
SQ   Sequence 447 BP; 116 A; 128 C; 112 G; 91 T; 0 U; 0 Other;

  Query Match             18.7%;  Score 130.8;  DB 14;  Length 447;
  Best Local Similarity   94.3%;  Pred. No. 3e-27;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        186 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 245

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        246 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 305

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        306 GGCAATCACCTGTAGCAGCA 325


RESULT 15
AAA35100
ID   AAA35100 standard; DNA; 562 BP.
XX
AC   AAA35100;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2789.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US017712.
XX
PR   03-AUG-1998;   98US-0095212P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstruction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers.
XX
PS   Disclosure; Page 1056; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an antisense
CC   oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC   nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC   impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of the
CC   ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC   sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC   from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC   AAA33992) are specifically claimed ONs from the present invention. N.B.
CC   Sequences given in the disclosure of the present invention do not match
CC   up with their corresponding SEQ ID NO: sequences given in the sequence
CC   listing
XX
SQ   Sequence 562 BP; 147 A; 163 C; 125 G; 127 T; 0 U; 0 Other;

  Query Match             18.7%;  Score 130.8;  DB 3;  Length 562;
  Best Local Similarity   94.3%;  Pred. No. 3.3e-27;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        223 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 282

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        283 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 342

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        343 GGCAATCACCTGTAGCAGCA 362


Search completed: February  9, 2006, 18:54:17
Job time : 454.558 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 15:02:28 ; Search time 2619.65 Seconds
(without alignments)
12519.885 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 699.8
Sequence: 1 gtcgtgtcactgctgtccag.....tgcagcctcaacctcctgag 701

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	181.2	25.9	729	10	CE721339	CE721339 tigr-gss-
c	2	156.8	22.4	653	10	CE574035	CE574035 tigr-gss-
c	3	130.8	18.7	425	1	AW139696	AW139696 UI-H-BI1-
c	4	130.8	18.7	459	1	AA845006	AA845006 ak61f09.s
	5	130.8	18.7	495	2	BG142111	BG142111 ia49b01.y
c	6	130.8	18.7	497	1	AI985509	AI985509 ws08d11.x
c	7	130.8	18.7	520	5	BU952425	BU952425 io74d05.x
	8	130.8	18.7	522	1	AW583679	AW583679 ia03a06.y
c	9	130.8	18.7	526	3	BQ271518	BQ271518 ik15a01.x
c	10	130.8	18.7	529	1	AA835293	AA835293 ak66b09.s
	11	130.8	18.7	529	3	BQ101169	BQ101169 ij24h05.y
c	12	130.8	18.7	532	1	AA835315	AA835315 ak66e08.s
	13	130.8	18.7	537	5	BU950978	BU950978 io74d05.y
c	14	130.8	18.7	539	1	AA844927	AA844927 ak61a06.s
	15	130.8	18.7	552	3	BM264109	BM264109 ig31a01.y
c	16	130.8	18.7	562	1	AI925956	AI925956 wh12g03.x
	17	130.8	18.7	572	3	BP323152	BP323152 BP323152
	18	130.8	18.7	577	4	BC013384	BC013384 Homo sapi
	19	130.8	18.7	579	3	BP322008	BP322008 BP322008
	20	130.8	18.7	580	3	BP323927	BP323927 BP323927
	21	130.8	18.7	581	3	BP324031	BP324031 BP324031
	22	130.8	18.7	581	3	BP324615	BP324615 BP324615
	23	130.8	18.7	582	3	BP323989	BP323989 BP323989
	24	130.8	18.7	582	3	BP324302	BP324302 BP324302
	25	130.8	18.7	582	3	BP325524	BP325524 BP325524
	26	130.8	18.7	583	3	BP323356	BP323356 BP323356
	27	130.8	18.7	584	3	BP322177	BP322177 BP322177
	28	130.8	18.7	584	3	BP322424	BP322424 BP322424
	29	130.8	18.7	584	3	BP322827	BP322827 BP322827
	30	130.8	18.7	584	3	BP322937	BP322937 BP322937
	31	130.8	18.7	584	3	BP325167	BP325167 BP325167
	32	130.8	18.7	643	1	AW951287	AW951287 EST363357
	33	130.8	18.7	742	2	BE969737	BE969737 601679687
	34	130.8	18.7	1002	2	BE969664	BE969664 601679602
c	35	129.8	18.5	289	1	AA844966	AA844966 ak57e12.s
	36	129.2	18.5	356	6	CA867922	CA867922 ir83a11.y
	37	129.2	18.5	380	5	BX113838	BX113838 BX113838
c	38	129.2	18.5	405	6	CA943648	CA943648 ir83a11.x
c	39	129.2	18.5	405	6	CA943649	CA943649 ir83a12.x
	40	129.2	18.5	416	6	CA867923	CA867923 ir83a12.y
	41	129.2	18.5	444	3	BQ126651	BQ126651 ii17e08.y
	42	129.2	18.5	459	5	BU949393	BU949393 in66f07.y
	43	129.2	18.5	466	5	BU071991	BU071991 im52h05.y
	44	129.2	18.5	479	3	BM505143	BM505143 ig90h06.y
	45	129.2	18.5	496	7	CK902774	CK902774 in68e07.x

ALIGNMENTS

RESULT 1
CE721339/c
LOCUS CE721339 729 bp DNA linear GSS 30-SEP-2003
DEFINITION tigr-gss-dog-17000329661665 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE721339
VERSION CE721339.1 GI:37061364
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 729)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .729
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Query Match 25.9%; Score 181.2; DB 10; Length 729;
Best Local Similarity 63.4%; Pred. No. 5.3e-31;
Matches 364; Conservative 3; Mismatches 181; Indels 26; Gaps 5;

Qy 34 GCAACTCTGTCTCAAAAAAAAAAAATGCTTTCAATAAATATATGATAAAAGGACTTATAT 93

```
Db        573 GAAAATCTTTTTTAAAAATACTAAATATAAATAAATAAACCTGATCAAAGGATTTAAGTT 514

Qy         94 TTTTTCAAGCCATAGGATC-ATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGT 152

Db        513 TTTTTTGAGCCACAGGACCAATAACTTCTGGAGCATGTCCACTAAATCATCCCTCCCTAC 454

Qy        153 TCCTGAGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTG 212

Db        453 CTATTCCTG-----------AGGCTGAGCTGCTTCCCTACCCCTTC----CTCAGCCCTG 409

Qy        213 TCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAG 272

Db        408 TTCTTCTCCCTCTCCAGGTGTTGCCAGACGCATGACCACTGCTACTCAGAAGCCAAGAAA 349

Qy        273 CTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGC 332

Db        348 CTGGACAGCTGTAAATTCCTCCTGGACAACCCCTACACCAAAATCTACTCATACTCATGC 289

Qy        333 TCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAG 392

Db        288 TCCGGCTCTGAGATCACCTGCAGCAGTAGGTTTAGCCCCTTGTTGGCTTGTCAGCCCTCA 229

Qy        393 TTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTAATT 452

Db        228 GTAGACAT---------GGGAGGAAAGAATAATAATGACATCCACCATTGATCAATCACA 178

Qy        453 TTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTACAA 512

Db        177 TCCTTGGTGCTAGGCTGTGTCTCCTTGAATCTTGGCAACAGCCCTAGGTGATCGGAATTA 118

Qy        513 TTATCCT-CACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAG 571

Db        117 TTATTATACCCATTTTACAGATGACAAAACAGCCTCAGAAGGTTGAGCCAATTGCCATGG 58

Qy        572 ATCACACAGCTTGTAAGTGGTGACAGTTTGGGTT 605

Db         57 GGCACACAGCTCGTAAATGGTGGTGTTTGGTTTT 24
```

RESULT 2
CE574035/c

```
LOCUS       CE574035                653 bp    DNA     linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-17000313102249 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE574035
VERSION     CE574035.1  GI:36890816
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN
```

Query Match 22.4%; Score 156.8; DB 10; Length 653;
Best Local Similarity 62.2%; Pred. No. 2.2e-25;
Matches 336; Conservative 3; Mismatches 175; Indels 26; Gaps 5;

```
Qy         34 GCAACTCTGTCTCAAAAAAAAAAAATGCTTTCAATAAATATATGATAAAAGGACTTATAT 93

Db        540 GAAAATCTTTTTTAAAAATACAAAATATAAATAAATAAACCTGATCAAAGGATTTAAGTT 481

Qy         94 TTTTTCAAGCCATAGGATC-ATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGT 152

Db        480 TTTTTTGAGCCACAGGACCAATAACTTCTGGAGCATGTCCACTAAATCATCCCTCCCTAC 421

Qy        153 TCCTGAGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTG 212

Db        420 CTATTCCTG-----------AGGCTGAGCTGCTTCCCTACCCCTTC----CTCAGCCCTG 376

Qy        213 TCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAG 272

Db        375 TTCTTCTCCCTCTCCAGGTGTTGCCAGACGCATGACCACTGCTACTCAGAAGCCAAGAAA 316

Qy        273 CTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGC 332

Db        315 CTGGACAGCTGTAAATTCCTCCTGGACAACCCCTACACCAAAATCTACTCATACTCATGC 256

Qy        333 TCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAG 392

Db        255 TCCGGCTCTGAGATCACCTGCAGCAGTAGGTTTAGCCCCTTGTTGGCTTGTCAGCCCTCA 196

Qy        393 TTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTA-GTGTTAAT 451

Db        195 GTAGACAT---------GGGAGGAAAGAATAATAATGACATCCACCATTGATCAATCACA 145

Qy        452 TTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTACA 511

Db        144 TCCCTTGGTGCTAGGCTGTGTCTCCTTGAATCTTGGCAACAGCCCTAGGTGATCGGAATT 85

Qy        512 ATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAG 571

Db         84 ATTATTATACCCATTTTACAGATGACAAAACAGCTCAGAAGGTTGAGCCAATTGACCATG 25
```

RESULT 3
AW139696/c

```
LOCUS       AW139696                425 bp    mRNA    linear   EST 30-OCT-1999
DEFINITION  UI-H-BI1-adr-h-03-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
            IMAGE:2717813 3', mRNA sequence.
ACCESSION   AW139696
VERSION     AW139696.1  GI:6144414
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .425
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2717813"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="NCI_CGAP_Sub3"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NCI_CGAP_Sub3 library is a subtracted library derived from
                      the NCI_CGAP_Sub1 library, which is a subtracted library
                      derived from BI. BI constitutes a mixture of 21
                      normalized or subtracted NCI_CGAP libraries:
                      NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
                      NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
                      NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
                      NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2,
                      NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
                      NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
                      NCI_CGAP_Brn25. These 21 libraries were pooled and a
                      single-stranded DNA preparation of the resulting mixture
                      was used as a tracer in a subtractive hybridization with
                      a driver whose composition is detailed below:
                      NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
                      3798-3803 (IMAGE CloneIDs 1322376-1323911,
                      1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
                      LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
                      1323912-1325831, 1471368-1472903, 1492104-1493255);
                      NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
                      CloneIDs 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
                      pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
                      CloneIDs 1257096-1258631, 1469064-1470983,
                      1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
                      2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
                      1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
                      LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,
                      1144584-1145351). Subtraction was performed as previously
                      described [Bonaldo, Lennon & Soares (1996): Normalization
                      and Subtraction: Two Approaches To Facilitate Gene
                      Discovery. Genome Research 6, 791-806.
                      TAG_TISSUE=lung
                      TAG_LIB=NCI_CGAP_Lu5
                      TAG_SEQ=CAAC"
ORIGIN

  Query Match            18.7%;  Score 130.8;  DB 1;  Length 425;
  Best Local Similarity  94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||:||||||||||||||||||||||||
Db        355 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 296

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        295 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 236

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        235 GGCAATCACCTGTAGCAGCA 216


RESULT 4
AA845006/c
LOCUS       AA845006                459 bp    mRNA    linear   EST 04-MAR-1998
DEFINITION  ak61f09.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
            IMAGE:1412393 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA845006
VERSION     AA845006.1  GI:2931457
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 377.
FEATURES              Location/Qualifiers
     source           1. .459
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1412393"
                      /sex="female"
                      /dev_stage="adult, 34 years"
                      /lab_host="DH10B"
                      /clone_lib="Barstead pancreas HPLRB1"
                      /note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)
                      with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
                      1st strand cDNA was primed with a Not I - oligo(dT) primer
                      [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I and Eco RI sites of the modified pT7T3 vector.
                      Library constructed by Bob Barstead. "
ORIGIN

  Query Match            18.7%;  Score 130.8;  DB 1;  Length 459;
  Best Local Similarity  94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||:||||||||||||||||||||||||
Db        353 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 294

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        293 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 234

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        233 GGCAATCACCTGTAGCAGCA 214


RESULT 5
BG142111
LOCUS       BG142111                495 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  ia49b01.y1 Melton Human Islets HIZ1 Homo sapiens cDNA clone
            IMAGE:5641656 5' similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG142111
VERSION     BG142111.1  GI:12645506
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,
            Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B.,
            Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,
            McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and
            Bowers,Y.
  TITLE     WashU-Harvard Pancreas EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5641656"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="TOP10"
                     /clone_lib="Melton Human Islets HIZ1"
                     /note="Organ: Pancreas; Vector: pZErO-2; Site_1: Not I;
                     Site_2: Xho I; Library constructed using SuperScript
                     Plasmid Library kit (Life Technologies). cDNA made by
                     oligo-dT priming. Xho I site destroyed during cloning.
                     Size-selected by column fractionation; average insert
                     size 1.59 kb. Primary library, unamplified."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 2;  Length 495;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        178 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 237

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        238 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 297

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        298 GGCAATCACCTGTAGCAGCA 317


RESULT 6
AI985509/c
LOCUS       AI985509                497 bp    mRNA    linear   EST 31-AUG-1999
DEFINITION  ws08d11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2496597 3'
            similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR (HUMAN);, mRNA
            sequence.
ACCESSION   AI985509
VERSION     AI985509.1  GI:5812786
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 386.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2496597"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid11"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 1;  Length 497;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        343 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 284

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        283 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 224

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        223 GGCAATCACCTGTAGCAGCA 204


RESULT 7
BU952425/c
LOCUS       BU952425                520 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  io74d05.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6132129 3'
            similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU952425
VERSION     BU952425.1  GI:24204177
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
```

```
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: io74d05.y1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6132129"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 5;  Length 520;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        335 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 276

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        275 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 216

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        215 GGCAATCACCTGTAGCAGCA 196


RESULT 8
AW583679
LOCUS       AW583679                 522 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  ia03a06.y1 Human Pancreatic Islets Homo sapiens cDNA clone
            IMAGE:5637202 5' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AW583679
VERSION     AW583679.1  GI:7260653
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D.,
            Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B.,
            Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,
            McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and
            Bowers,Y.
  TITLE     WashU-Harvard Pancreas EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ia03a06.x3
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Libraries were constructed by Dr. Douglas Melton
             DNA sequencing by: Washington University Genome Sequencing Center
            For information on obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
             This sequence now available from the IMAGE consortium, for clone
            orders contact: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5637202"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="Human Pancreatic Islets"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Library constructed using SuperScript
                     Plasmid Library kit (Life Technologies). cDNA made by
                     oligo-dT priming.  Size-selected by column fractionation;
                     average insert size 1.08 kb. Primary library,
                     unamplified."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 1;  Length 522;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        192 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 251

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db        252 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 311

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        312 GGCAATCACCTGTAGCAGCA 331


RESULT 9
BQ271518/c
LOCUS       BQ271518                 526 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  ik15a01.x1 HR85 islet Homo sapiens cDNA clone IMAGE:5781000 3'
            similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BQ271518
VERSION     BQ271518.1  GI:20496584
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 526)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40UP from Gibco
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .526
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5781000"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match            18.7%;  Score 130.8;  DB 3;  Length 526;
  Best Local Similarity  94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        341 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 282

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::||||||||||||||||||||||||||||||||||||||
Db        281 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 222

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        221 GGCAATCACCTGTAGCAGCA 202


RESULT 10
AA835293/c
LOCUS       AA835293                529 bp    mRNA    linear   EST 23-FEB-1998
DEFINITION  ak66b09.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
            IMAGE:1412825 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA835293
VERSION     AA835293.1  GI:2909021
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1412825"
                     /sex="female"
                     /dev_stage="adult, 34 years"
                     /lab_host="DH10B"
                     /clone_lib="Barstead pancreas HPLRB1"
                     /note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. "
ORIGIN

  Query Match            18.7%;  Score 130.8;  DB 1;  Length 529;
  Best Local Similarity  94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||
Db        342 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 283

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::||||||||||||||||||||||||||||||||||||||
Db        282 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 223

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        222 GGCAATCACCTGTAGCAGCA 203


RESULT 11
BQ101169
LOCUS       BQ101169                529 bp    mRNA    linear   EST 29-APR-2002
DEFINITION  ij24h05.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:6135800 5' similar to SW:PA21_HUMAN P04054
            PHOSPHOLIPASE A2 PRECURSOR ;, mRNA sequence.
ACCESSION   BQ101169
VERSION     BQ101169.1  GI:20134153
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
```

```
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ij24h05.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center This clone is
            available royalty-free through LLNL; please contact the IMAGE
            consortium (info@image.llnl.gov) for further information
            Putative full length read
            vector to vector length is 572
            Seq primer: -40RP from Gibco
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6135800"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 3;  Length 529;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||:|||||||||||||||||||||||||
Db        190 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 249

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||
Db        250 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 309

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        310 GGCAATCACCTGTAGCAGCA 329


RESULT 12
AA835315/c
LOCUS       AA835315                532 bp    mRNA    linear   EST 23-FEB-1998
DEFINITION  ak66e08.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
            IMAGE:1412870 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA835315
VERSION     AA835315.1  GI:2909043
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1412870"
                     /sex="female"
                     /dev_stage="adult, 34 years"
                     /lab_host="DH10B"
                     /clone_lib="Barstead pancreas HPLRB1"
                     /note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. "
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 1;  Length 532;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||:|||||||||||||||||||||||||
Db        344 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 285

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||
Db        284 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 225

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        224 GGCAATCACCTGTAGCAGCA 205


RESULT 13
BU950978
LOCUS       BU950978                537 bp    mRNA    linear   EST 21-OCT-2002
DEFINITION  io74d05.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6132129 5'
            similar to SW:PA21_HUMAN P04054 PHOSPHOLIPASE A2 PRECURSOR ;, mRNA
            sequence.
ACCESSION   BU950978
VERSION     BU950978.1  GI:24202730
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Putative full length read
            vector to vector length is 572
            Seq primer: -40RP from Gibco
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6132129"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 5;  Length 537;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        223 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 282

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::||||||||||||||||||||||||||||||||||||||||
Db        283 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 342

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        343 GGCAATCACCTGTAGCAGCA 362


RESULT 14
AA844927/c
LOCUS       AA844927                539 bp    mRNA    linear   EST 04-MAR-1998
DEFINITION  ak61a06.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
            IMAGE:1412338 3' similar to gb:M21054 PHOSPHOLIPASE A2 PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA844927
VERSION     AA844927.1  GI:2931378
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1412338"
                     /sex="female"
                     /dev_stage="adult, 34 years"
                     /lab_host="DH10B"
                     /clone_lib="Barstead pancreas HPLRB1"
                     /note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. "
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 1;  Length 539;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || |||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        350 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 291

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::||||||||||||||||||||||||||||||||||||||||
Db        290 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 231

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        230 GGCAATCACCTGTAGCAGCA 211


RESULT 15
BM264109
LOCUS       BM264109                552 bp    mRNA    linear   EST 18-DEC-2001
DEFINITION  ig31a01.y1 HR85 islet Homo sapiens cDNA 5' similar to SW:PA21_HUMAN
            P04054 PHOSPHOLIPASE A2 PRECURSOR ;, mRNA sequence.
ACCESSION   BM264109
VERSION     BM264109.1  GI:17927149
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match             18.7%;  Score 130.8;  DB 3;  Length 552;
  Best Local Similarity   94.3%;  Pred. No. 2.1e-19;
  Matches  132;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        221 ACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAG 280
              |||   || ||||||||||||||||||||||||:||||||||||||||||||||||
Db        201 ACTGGACAAGTGCTGCCAGACACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAG 260

Qy        281 CTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 340
              ||||||||||||||||||::||||||||||||||||||||||||||||||||||||
Db        261 CTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTC 320

Qy        341 GGCAATCACCTGTAGCAGTA 360
              |||||||||||||||||| |
Db        321 GGCAATCACCTGTAGCAGCA 340


Search completed: February 10, 2006, 00:58:05
Job time : 2623.65 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:19:56 ; Search time 1611.05 Seconds
(without alignments)
10346.286 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 2499.4
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2499.4	100.0	12174	12	ADJ09983	Adj09983 Human pho
→	2	2499.4	100.0	13612	6	ABK47376	Abk47376 Human Pho
	3	2498.6	100.0	12174	12	ADI35082	Adi35082 Human PLA
c	4	457	18.3	42360	13	ABD33466	Abd33466 Human can
	5	434.2	17.4	93544	13	ABD33504	Abd33504 Human can
c	6	434.2	17.4	160482	11	ACN43914	Acn43914 Human gen
c	7	433.2	17.3	91760	11	ACN44410	Acn44410 Human gen
c	8	425.4	17.0	38538	13	ABD33357	Abd33357 Human can
	9	417.2	16.7	122888	6	ABK83569	Abk83569 Human cDN
c	10	416.2	16.7	6405	5	AAF97850	Aaf97850 Human neu
c	11	416	16.6	126990	12	ADP13332	Adp13332 Renal cel
c	12	414.2	16.6	196686	11	ACN44170	Acn44170 Human gen
c	13	413.6	16.5	86000	12	ADP68568	Adp68568 Human PPA
c	14	413.6	16.5	215974	12	ADQ97523	Adq97523 Human can
c	15	413.4	16.5	13670	6	AAL42369	Aal42369 Human Gua
	16	413.2	16.5	6519	5	ABA16909	Aba16909 Human ner
	17	408.8	16.4	17245	4	AAK83897	Aak83897 Human imm
	18	406.2	16.3	348101	12	ADQ97146	Adq97146 Human can
c	19	405.8	16.2	227246	13	ABD33272	Abd33272 Human can
c	20	404	16.2	95240	10	ADL13556	Adl13556 Osteoarth
c	21	403.6	16.1	23639	6	ABQ78991	Abq78991 Human pho
c	22	400.4	16.0	14176	4	AAS26670	Aas26670 Human gen
c	23	400.4	16.0	14176	8	ABX74019	Abx74019 Human nov
c	24	400.4	16.0	177531	8	ACF62732	Acf62732 Cancer ba
c	25	400.4	16.0	177531	8	ADB20847	Adb20847 MRP1 base
c	26	400.4	16.0	177531	10	ADB87936	Adb87936 Human UGT
c	27	400.4	16.0	177531	10	ADB96919	Adb96919 Human MDR
c	28	400.4	16.0	177531	10	ADB92110	Adb92110 Human MDR
c	29	400.4	16.0	177531	10	ADH74617	Adh74617 Human BAC
	30	399	16.0	87687	11	ACN45166	Acn45166 Human gen
	31	399	16.0	131078	14	ADX06911	Adx06911 Cyclin-de
	32	398.6	15.9	174448	11	ACN43946	Acn43946 Human gen
c	33	398.4	15.9	167343	6	ABL64403	Abl64403 Stomach c
c	34	398.4	15.9	167343	6	ABL67239	Abl67239 Thyroid c
	35	394.8	15.8	30620	4	AAK66931	Aak66931 Human imm
	36	394.6	15.8	19300	12	ADP74371	Adp74371 Human X c
	37	392.4	15.7	9289	10	ADE84033	Ade84033 5' regula
	38	391.4	15.7	226475	9	AAD58279	Aad58279 Human tum
c	39	390	15.6	13224	4	AAS41751	Aas41751 Genomic s
c	40	390	15.6	13224	4	ABA06811	Aba06811 Human gen
c	41	390	15.6	13224	6	ABV84148	Abv84148 Human pol
	42	390	15.6	36221	4	AAS00624	Aas00624 Human dea
c	43	389.6	15.6	23456	13	ABD33110	Abd33110 Human can
	44	389.6	15.6	60057	11	ACN44314	Acn44314 Human gen
	45	388	15.5	75252	11	ACN44450	Acn44450 Human gen

ALIGNMENTS

RESULT 1
ADJ09983
ID ADJ09983 standard; DNA; 12174 BP.
XX
AC ADJ09983;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(436,c)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(839,a)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4050,a)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(4689,t)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6282,a)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6358,c)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(6653,t)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphisms"
FT variation replace(7256,t)
FT /*tag= h

```
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7300,a)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7301,c)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7328,g)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(8062,c)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(9182,t)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(11649,c)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphisms"
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   P-PSDB; ADJ09984.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ   Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2499.4;  DB 12;  Length 12174;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2497;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7000 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 7059

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7060 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 7119

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7120 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 7179

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7180 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       7240 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 7299

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7300 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7360 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 7419

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7420 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 7479

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7480 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7539

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7540 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7599

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7600 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7659

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7660 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7719

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7720 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7779

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7780 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7839

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7840 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7899

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7900 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7959

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7960 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 8019

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       8020 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 8079

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8080 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 8139

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8140 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 8199

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
```

```
Db       8200 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 8259

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db       8260 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 8319

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTTCTT 1380
Db       8320 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTTCTT 8379

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db       8380 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 8439

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
Db       8440 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8499

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
Db       8500 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8559

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
Db       8560 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8619

Qy       1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
Db       8620 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8679

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
Db       8680 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8739

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
Db       8740 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8799

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
Db       8800 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8859

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
Db       8860 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8919

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
Db       8920 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8979

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
Db       8980 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 9039

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
Db       9040 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 9099

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
Db       9100 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 9159

Qy       2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
Db       9160 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 9219

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
Db       9220 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 9279

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
Db       9280 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 9339

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
Db       9340 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 9399

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
Db       9400 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 9459

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
Db       9460 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9500
```

RESULT 2
ABK47376
ID ABK47376 standard; DNA; 13612 BP.
XX
AC ABK47376;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(3845,A)
FT		/*tag= a
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	variation	replace(3968,A)
FT		/*tag= b
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	CDS	4053. .9613
FT		/*tag= c
FT		/product= "Pla2G1B"
FT	exon	4053. .4086
FT		/*tag= d
FT		/number= 1
FT	intron	4087. .5785
FT		/*tag= e
FT		/number= 1
FT	exon	5786. .5945
FT		/*tag= f
FT		/number= 2
FT	intron	5946. .6744
FT		/*tag= g
FT		/number= 2
FT	variation	replace(6060,A)
FT		/*tag= h
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	exon	6745. .6873
FT		/*tag= i
FT		/number= 3
FT	variation	replace(6844,A)
FT		/*tag= j
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	intron	6874. .9489
FT		/*tag= k
FT		/number= 3
FT	exon	9490. .9613
FT		/*tag= l
FT		/number= 4
FT	variation	replace(9531,A)

```
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 2499.4;  DB 6;  Length 13612;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2497;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6816 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7116 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7235

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7236 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7295

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7296 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7355

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7356 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7415

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7416 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7475

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7476 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 7535

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db       7536 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 7595

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7596 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 7655

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7656 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 7715

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7716 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 7775

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
```

```
Db       7776 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 7835

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7836 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 7895

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7896 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 7955

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7956 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8015

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8016 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8075

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8076 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8135

Qy       1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8136 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8195

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8196 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8255

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8256 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8315

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8316 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8375

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8376 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8435

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8436 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8495

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8496 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 8555

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8556 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 8615

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8616 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 8675

Qy       2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8676 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 8735

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8736 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 8795

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8796 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 8855

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       8856 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 8915

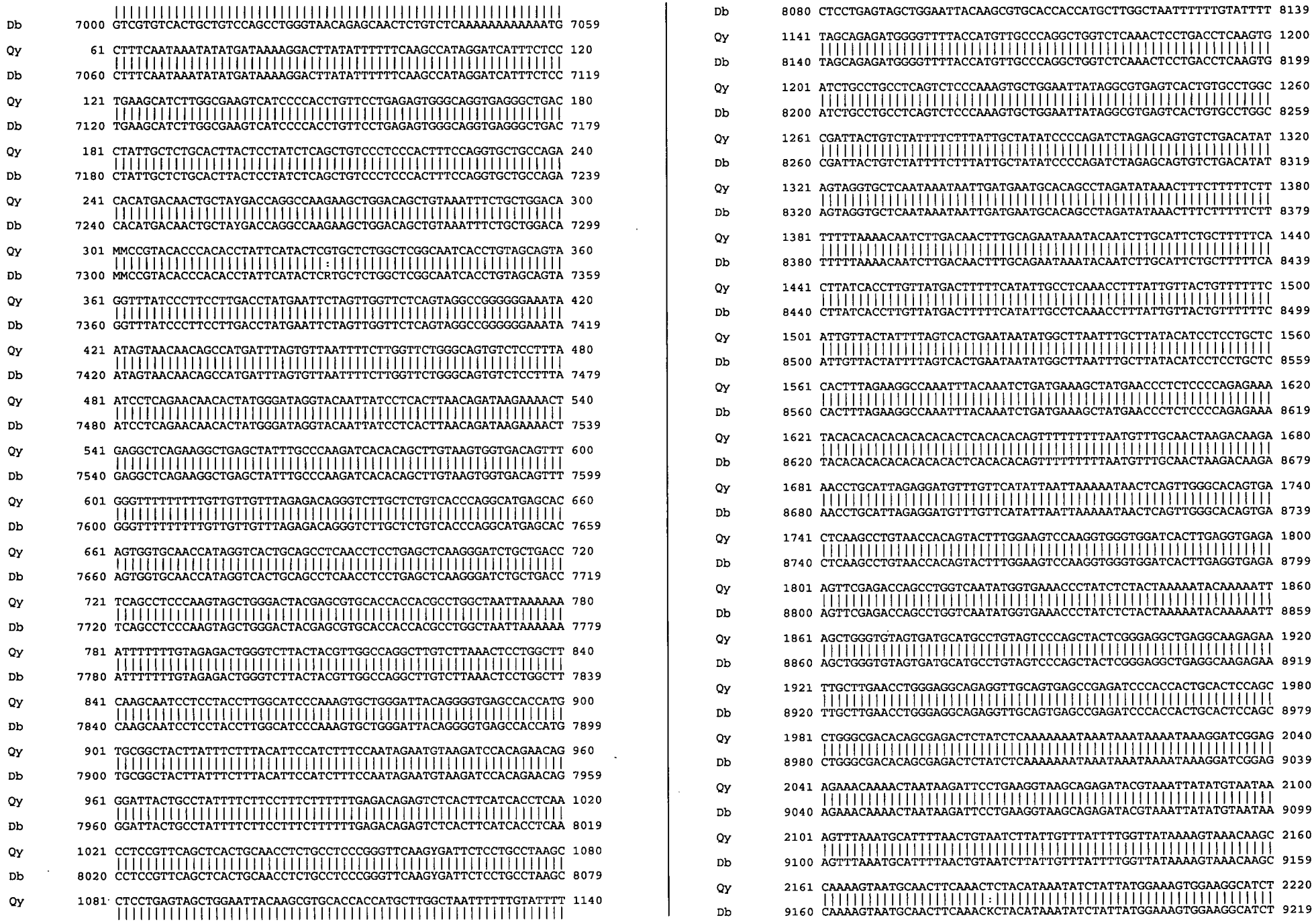
Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8916 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 8975

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
              |||||||||||||||||||||||||||||||||||||||||
Db       8976 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9016
```

```
RESULT 3
ADI35082
ID   ADI35082 standard; DNA; 12174 BP.
XX
AC   ADI35082;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human PLA2G1B nucleotide sequence.
XX
KW   PLA2G1B ; fat deposition; leanness; polymorphism;
KW   non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW   hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW   microalbuminuria; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004002295-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020830.
XX
PR   27-JUN-2002; 2002US-0392361P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML;
XX
DR   WPI; 2004-082843/08.
DR   P-PSDB; ADI35083.
XX
PT   Diagnosing a predisposition to fat deposition or leanness, useful for
PT   diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT   detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT   the subject.
XX
PS   Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC   The invention relates to diagnosing a predisposition to fat deposition or
CC   leanness in a subject comprising detecting the presence or absence of a
CC   polymorphic variation associated with fat deposition at a polymorphic
CC   site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC   subject, where the presence of the polymorphic variation indicates a
CC   predisposition to fat deposition in the subject. The polymorphic
CC   variation is a guanine at position 7328 or thymine at position 9182 of
CC   the present sequence. The method is useful for diagnosing a
CC   predisposition to fat deposition or leanness in a subject, and
CC   consequently for diagnosing a predisposition to non-insulin dependent
CC   diabetes mellitus (NIDDM) in a subject and conditions such as
CC   hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC   hypercoagulability, or microalbuminuria, which can lead to early
CC   prescription of preventive measures. The present sequence represents a
CC   human PLA2G1B nucleotide sequence.
XX
SQ   Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 2499.4;  DB 12;  Length 12174;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2499;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
```


Db 7000 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 7059
Qy 61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
Db 7060 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 7119
Qy 121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
Db 7120 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 7179
Qy 181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
Db 7180 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239
Qy 241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
Db 7240 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 7299
Qy 301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
Db 7300 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359
Qy 361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
Db 7360 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 7419
Qy 421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
Db 7420 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 7479
Qy 481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
Db 7480 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7539
Qy 541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
Db 7540 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7599
Qy 601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
Db 7600 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7659
Qy 661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
Db 7660 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7719
Qy 721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
Db 7720 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7779
Qy 781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
Db 7780 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7839
Qy 841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
Db 7840 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7899
Qy 901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
Db 7900 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7959
Qy 961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
Db 7960 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 8019
Qy 1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db 8020 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 8079
Qy 1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db 8080 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 8139
Qy 1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db 8140 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 8199
Qy 1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db 8200 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 8259
Qy 1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db 8260 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 8319
Qy 1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db 8320 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 8379
Qy 1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db 8380 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 8439
Qy 1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
Db 8440 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8499
Qy 1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
Db 8500 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8559
Qy 1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
Db 8560 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8619
Qy 1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
Db 8620 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8679
Qy 1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
Db 8680 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8739
Qy 1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
Db 8740 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8799
Qy 1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
Db 8800 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8859
Qy 1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
Db 8860 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8919
Qy 1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
Db 8920 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8979
Qy 1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
Db 8980 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 9039
Qy 2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
Db 9040 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 9099
Qy 2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
Db 9100 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 9159
Qy 2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
Db 9160 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 9219

```
Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9220 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 9279

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9280 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 9339

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9340 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 9399

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9400 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 9459

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db       9460 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9500
```

RESULT 4
ABD33466/c
ID ABD33466 standard; DNA; 42360 BP.
XX
AC ABD33466;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-089.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 618; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 42360 BP; 11698 A; 9658 C; 10445 G; 10559 T; 0 U; 0 Other;

Query Match 18.3%; Score 457; DB 13; Length 42360;
Best Local Similarity 100.0%; Pred. No. 4.3e-84;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       2045 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 2104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42360 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 42301

Qy       2105 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 2164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42300 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 42241

Qy       2165 AGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 2224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42240 AGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 42181

Qy       2225 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42180 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 42121

Qy       2285 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 2344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42120 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 42061

Qy       2345 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 2404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42060 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 42001

Qy       2405 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 2464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42000 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 41941

Qy       2465 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||
Db      41940 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 41904
```

RESULT 5
ABD33504
ID ABD33504 standard; DNA; 93544 BP.
XX
AC ABD33504;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-098.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 676; 182pp; English.
XX

```
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 434.2;  DB 13;  Length 93544;
  Best Local Similarity   59.7%;  Pred. No. 2.6e-79;
  Matches  861;  Conservative    1;  Mismatches  559;  Indels   22;  Gaps    7;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
Db      18099 TTTTTTTTTTTTTTTTTTGAGAGATGGAGTCTTAGCCTGTTGCCCAGGCTGGAGTGCAGT 18158

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
Db      18159 GGCACCATCTTGGGTCACTGCAACCTCCACCATGTGGGTTCAAGAGATTCTACTGCCTCA 18218

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db      18219 ACCTCCCAAGTAGCTGGAATTACAGGCTTGTGCCACCACACCTTGCTAATT--TTTGTGT 18276

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db      18277 TTTTATTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGATCTCAAACTCCTGGCCTCAA 18336

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db      18337 CTGATCCGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCACCATGCCT 18396

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db      18397 GGCCACCACAGGCTTTATTTACAGGTTGCCTCATTACCTGAATTCTCTCTCTCCAAATTC 18456

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA---A 1020
Db      18457 TTGTTACCCAAACACAGGAACATGAGATGGAGTCTCGCTCCGTCGCCCAGTCTGGAGTGC 18516

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db      18517 AGTGATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTTCTGCCTCAGC 18576

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db      18577 CTCCCAAGTAGCTGGGACTACAGGCATGCGCCACCCCACCTGGCTAA-TTTTTGTATTTT 18635

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db      18636 TAGTAGAGACGGGGTTTCACCATGCTGGCCAGGCTGGTCTCGAACTCCTGACCTC--GTG 18693

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db      18694 ATCCACCCGCCTCAGCTTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATACCTGGC 18753

Qy       1261 CGATTACTGTCTATTTT------CTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314
Db      18754 CAGGAACTAGACAGTTTTAGAGACATCTTTCTGAAATAGGGCTGGTGCCCCTCATTATTC 18813

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374
Db      18814 ATTAATGCTATTTTCTCCAAAACCCGGGAACCAAATAGATGTGATAGCATGGCATCCCTC 18873

Qy       1375 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434
Db      18874 ATGATCTGAAGTTACTGTCCGTAACTGCGGTCCCATTCGATTTGGATATTGGGGGATGCC 18933

Qy       1435 TTTTCACTTATCA-CCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTG 1493
Db      18934 TGCACGTGTCTGTGTCATGTGCTGCATGTCTGCGGTGCCATCAGTCTGTGAGCTCCCTCT 18993

Qy       1494 TTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCT 1553
Db      18994 GGGCCTGCGTCCGTTTTTCTCATCCTTTGTATTATCTCAGCACTTTGTACAATACCTGGC 19053

Qy       1554 CCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCC 1613
Db      19054 ACTCAATGAATTAAAAATAATCTGAAGCGGGGGTTGGGGGGAGCTTGTCAATATAGCCCC 19113

Qy       1614 AGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAA 1673
Db      19114 ACCTTGTTTTGCCCTTGGACTTGATAAAGCCGTTTGTTGCTATGCATATGTCCATCTCTT 19173

Qy       1674 GACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCA------- 1726
Db      19174 TTCTAGACAATGAGTTACTTGGATGAATAGGCAGATCTCTTGGTTAAACCATAAAGACAG 19233

Qy       1727 GTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGAT 1786
Db      19234 GCTGGGCGCGGTGGCTCACATCTGTAATCCCAGCACTTTGAGAGGCCGAGGCGGGCAGAT 19293

Qy       1787 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1846
Db      19294 CACTTGAGGTTAGAAGTTCGATACCAGCCTGGCCAACATGGTGAAACCCAATCTCTACTA 19353

Qy       1847 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1906
Db      19354 AAAATACAAAAATTAGCCGAGTGCGGTGGCGCATGCCTGTAATCCTAGCTACTTGGGAGG 19413

Qy       1907 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1966
Db      19414 CTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGAAGCTTTCAGTGAGCCGAGATTGTGC 19473

Qy       1967 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 2026
Db      19474 CATTACACTCCAGCCTGGGCAACAGAGCGAGACTGTGTCTCAAAATGAAAACCCATAAAG 19533

Qy       2027 ATA 2029
Db      19534 ACA 19536
```

```
RESULT 6
ACN43914/c
ID   ACN43914 standard; DNA; 160482 BP.
XX
AC   ACN43914;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG26773.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
```

```
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 100; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 160482 BP; 44060 A; 32143 C; 33530 G; 49875 T; 0 U; 874 Other;

  Query Match             17.4%;  Score 434.2;  DB 11;  Length 160482;
  Best Local Similarity   60.2%;  Pred. No. 2.9e-79;
  Matches  888;  Conservative    1;  Mismatches  525;  Indels   61;  Gaps    8;

Qy        606 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 665
Db       5181 TGTTTTTTTTTTTTCTCTCGATACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 5122

Qy        666 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 725
Db       5121 CATGATCTCGGCTCACTGAAACCTCTGCCTCCTGGGTTCACGTGATTCTCCTGCCTCAGA 5062

Qy        726 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGG--CTAATTAAAAAAATT 783
Db       5061 CTCCCAAGTAGCTGGGATTACAGGCGTGAGCCACCATGCCTGGCCCTAATGTTTTGTATT 5002

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db       5001 TTTTAGTAGAGAC-AGGTTTCGCTACATTGGTCAGGCTAGTCTCAAACTCCTGACCTCAA 4943

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db       4942 GTGATCCGCCTGCCTTGATGTACCAAAGTGCTGGGATTACAGATGTGAGCCACCCCGCCC 4883

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db       4882 GGCCAGTCATTCACTTTCTTTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCATTCTGTC 4823

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
Db       4822 ATCCAGGCTGGAGTGT----------------------------------AGTGGCAC 4799

Qy       1024 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 1083
Db       4798 GATCTCGGCTCACTTCAACTTTCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTC 4739

Qy       1084 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAG 1143
Db       4738 TTGAGTAGCTGGGATTACAGGCATGCACCACCGCACCCGGCTAATTTTTTATATTTTTGG 4679

Qy       1144 CAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATC 1203
Db       4678 GAGAGATGGGGTTTCACCATGTTAGACAGGCTGGTCTCAAACTCCTGACCTCAAGCAATC 4619

Qy       1204 TGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGA 1263
Db       4618 TGCTCGCCTTGGCCTCCTAAAGTGCTGGGATTACAGGCGTGAGCCACCACACCCGGCCAA 4559

Qy       1264 TTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGT 1323
Db       4558 --ATCATTCACTTTCTACAACAAGTTCTGTCTGTATGCCAATGAATCTCAGGACTGTACT 4501

Qy       1324 AGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTT 1383
Db       4500 TCTAGCTCTAACCTCTAAACATATATTTAAGTGTCCTTTAAAGCCAGGACTTAGATTCCT 4441

Qy       1384 TTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTT 1443
Db       4440 CAGAATCTCAAATTCAACAAATCCAAAATTCAACATATATTTTCCCTAAATTCTGCCCCT 4381

Qy       1444 ATCACCTTGTTA------TGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTT 1497
Db       4380 TTCTCTGTATTTCCTAGTTTGCTTAATAGTATCACCTTAATTCAAGGAAGAAATTTCAAA 4321

Qy       1498 TTCATTGTTACTATTTTAGTCACTGAAT--AATATGGCTTAATTTGCTTATACATCCTCC 1555
Db       4320 GTCATCTTCAATTTTTCTCTCTCTCAACGCAATAAAGCACTAATTCCTAGAAGCTCAACC 4261

Qy       1556 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1615
Db       4260 TTAAACAGTCCTCTGGAATGAGATGAAGGTTAACAGTTACTGAACACAGACCAAAACTAG 4201

Qy       1616 AGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGA 1675
Db       4200 GCACAGCAGCAAGGGTTGTTCAATTGAATTCAATTCTCTCAGAAGCCCTAAGGAATAGCT 4141

Qy       1676 CAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAAT------AACTCAGTT 1729
Db       4140 CCAATTTTATAGAGGAAGAAACTGAGGTTCCAAGAAGTAAGAAATTCACACAAGGAGGCT 4081

Qy       1730 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 1789
Db       4080 GGGTGTGGTGGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCCAAGGCGGGCAGATCAC 4021

Qy       1790 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 1849
Db       4020 CTGAGGTCAAGAGCTCGAGACCAGCCTGACAAACATGGTGAAACCCCATCTCTACTAAAA 3961

Qy       1850 ATACAAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 1909
Db       3960 ATACAAAAATTACCCAGGTGTGGTGGTGCACGCCTGTAATCCCAGCTATTTGGGAGGCTG 3901

Qy       1910 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCC------GAGATCC 1963
Db       3900 ACGCAGGAGAATTGTTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATGGAGATCG 3841

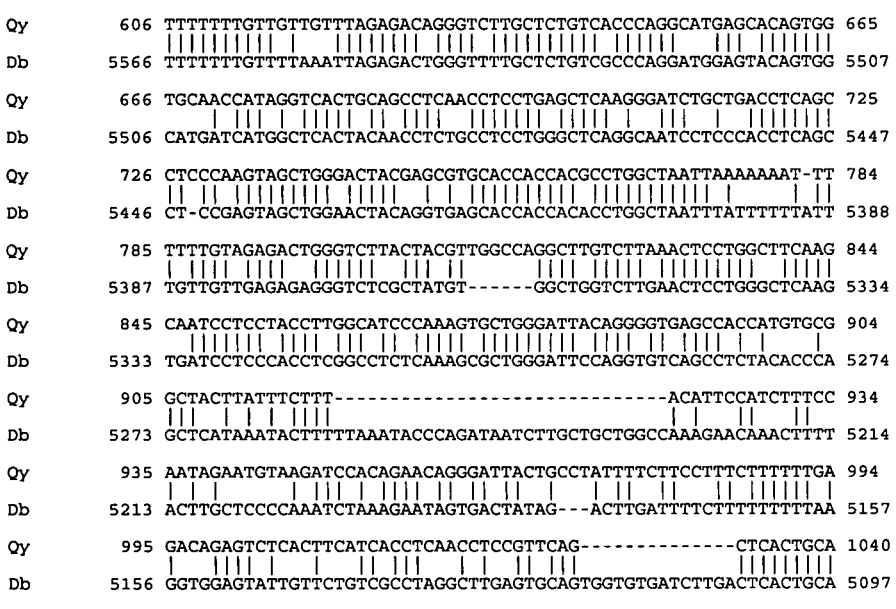
Qy       1964 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 2023
Db       3840 TGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCTGTCTCCAAAAAAAAAAAAAAA 3781

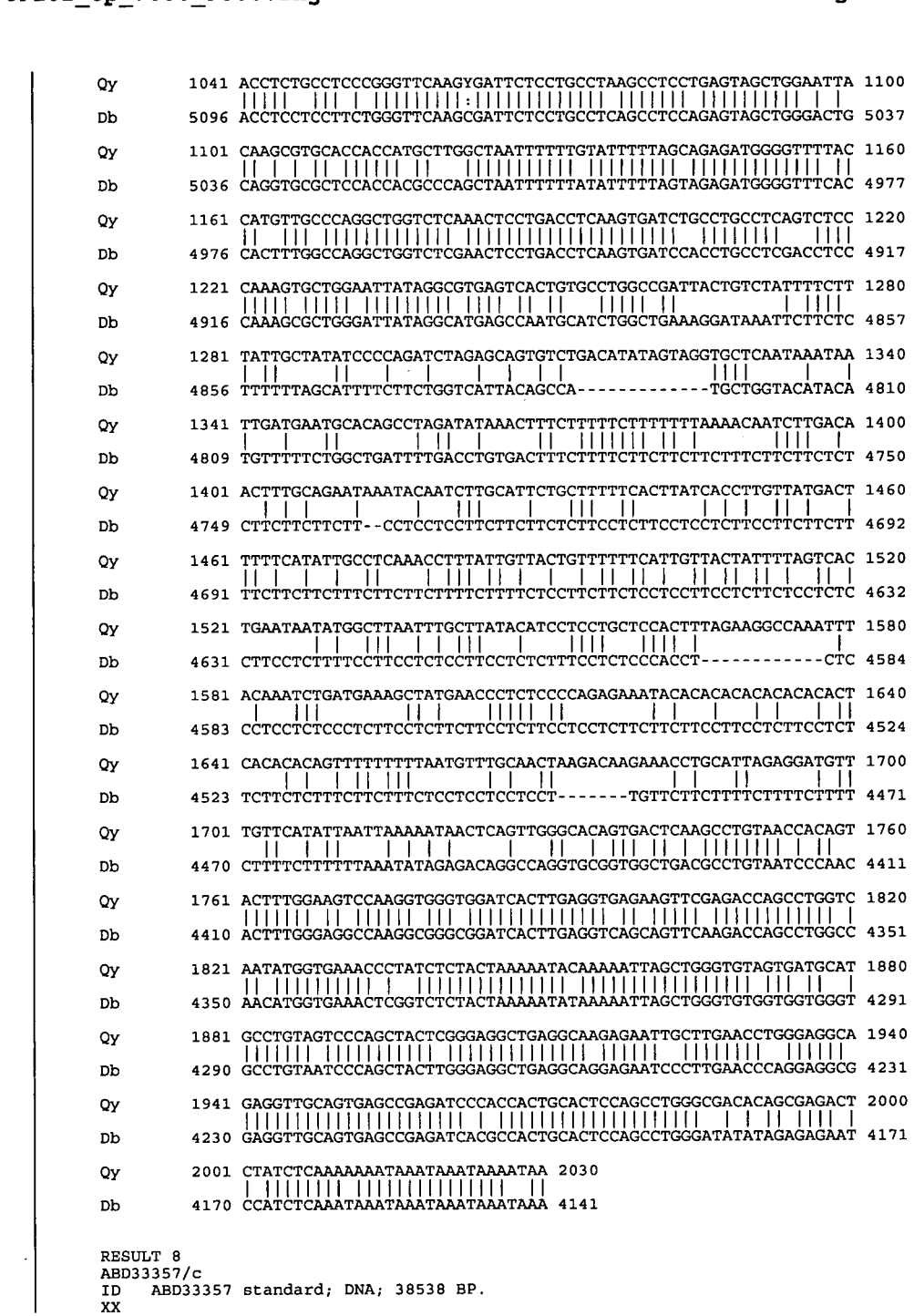
Qy       2024 AAAATAAAGGATCGGAGAGAAACAAAACTAATAAG 2058
Db       3780 AGAAAGAAAGAAAGAAAAAAATAAATTCACACAAG 3746

RESULT 7
ACN44410/c
ID   ACN44410 standard; DNA; 91760 BP.
XX
AC   ACN44410;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG41365.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
```

XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 844; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 91760 BP; 21177 A; 23589 C; 24129 G; 22845 T; 0 U; 20 Other;

Query Match 17.3%; Score 433.2; DB 11; Length 91760;
Best Local Similarity 61.6%; Pred. No. 4.1e-79;
Matches 906; Conservative 1; Mismatches 474; Indels 89; Gaps 10;





RESULT 8
ABD33357/c
ID ABD33357 standard; DNA; 38538 BP.
XX

AC ABD33357;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-065.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 448; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 38538 BP; 9418 A; 9893 C; 9523 G; 9684 T; 0 U; 20 Other;

Query Match 17.0%; Score 425.4; DB 13; Length 38538;
Best Local Similarity 60.0%; Pred. No. 1.4e-77;
Matches 879; Conservative 1; Mismatches 552; Indels 33; Gaps 9;

```
Qy      567 CCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAG 626
Db    35304 CAAAAAATAAAAATTAAATATTTTTTGTTTCTTAGATTATAGTTAAGGGGCTTTTTTGAG 35245

Qy      627 ACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCA 685
Db    35244 ATGAAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAATGGTGTGATCTTGGCTCACAGCA 35185

Qy      686 GCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTA 745
Db    35184 ACCTCCACCTCCAGGGTTCAAGCGATTCTCCTCTCTCAGCCTCCTGAGTAGCTGGGATTA 35125

Qy      746 CGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTA 805
Db    35124 CAGGCATGCGCCACCACACCCAGCTAATTTTGTA---TTTTTAGTAGAGATGGGGTTTCT 35068

Qy      806 CTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATC 865
Db    35067 CCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCACCCTCCTTGGCTTC 35008

Qy      866 CCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC 925
Db    35007 CCGAAGTGTTGGGATTACAGGCATGAGCCACTGCGCCTGGC------CTCAGTTAAGGGA 34954

Qy      926 CATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTT 985
Db    34953 CTTTTTACATATTTATTTATTTATTTATTATTATTTTGAGACGAAGTCTCGCTGTT 34894

Qy      986 CTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTC 1045
Db    34893 GTTCTCCAGGCTGGAGTGCAATGGCACGACC----------TTGGCTCACTGCAACCTC 34845

Qy     1046 TGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGC 1105
Db    34844 CGCCTCCCGGGTTCAAGCGATTCTCCTGTCTTGGCCTCCCAAGTAGCTGGGATTACAGGC 34785

Qy     1106 GTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGT 1165
Db    34784 GCCTGCCATCATGCCTGGCTAA-TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATTA 34726

Qy     1166 TGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAG 1225
Db    34725 TGGCCAGGCTGGTCTTGAAGTCCTGACTTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAG 34666

Qy     1226 TGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTG 1285
Db    34665 TGCTGGGATGACAGGTGTGAGCCACCACACTCGGCCTTTATTTATTATTTTCGAGACAGG 34606

Qy     1286 CTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGAT 1345
Db    34605 GTATCTC------TCTGTTGCCCAGGCTGGAGTGCAGTGGTGCAGTCACGGCTGGTTGCA 34552

Qy     1346 GAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTT 1405
Db    34551 GCCTCAAACTCCTGGGCTCCAGCA--ATTCTCCTGCCTCAGCCTCCTGGGTAGCTGGAAC 34494

Qy     1406 GCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTC 1465
Db    34493 TACAGGTGCAAATTATCATGCCTGGCTAATTTTTAAATTTTTTGTCAAGATGGGGTCTCT 34434

Qy     1466 ATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAAT 1525
Db    34433 CTATATTGCCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATCTCCCTTCTGGGCTTCA 34374

Qy     1526 AATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAA 1585
Db    34373 AAAAGTGCTGGGATTATAAGTATGAGCCACCACCTCCGGCCTTCCCTCCTTCCTTCCTTC 34314

Qy     1586 TCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACTCACAC 1645
Db    34313 CTTTCTGCCTTCCTTCCTTCCTTTCTCAAAAAAAAAAAAAAAAGAAGTCAGTGTTGTCTGA 34254

Qy     1646 ACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTC 1705
Db    34253 CTGGTATTGTCTGATTACCTGCATATATTGATATATTATTCCAATAAAGATGCTTGCTTT 34194

Qy     1706 ATA--TTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 1763
Db    34193 TTCCTTTGGTTAAAAAGTATCCGGCCAGGCGCGATGGCTCACGCTTGTAACCCCAGTACT 34134

Qy     1764 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 1823
Db    34133 TTGGGAGGCTGTGGTGGGCGGATCACCTGAGATCAGGAGTACGAGACCAGCCTGGCCAAC 34074

Qy     1824 ATGGTGAAACCCTATCTCTACTAAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 1883
Db    34073 ATGGTAAAACCCCATCTCTACTAGAAATACAAAAATTAGCTGAGTGTAGTGGCACGCACC 34014

Qy     1884 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 1943
Db    34013 TGTAATCCCAGCTACTTGGGAGGCCAAGGGAGAAGAATTGTTTGAACCCGGGAGGCAGAG 33954

Qy     1944 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 2003
```

```
                ||||||||||||| ||||| ||||| |||||||||||| |||||||| || ||||||||
Db      33953 GTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCAGGGCGACA-AGAGAGACTCTG 33895

Qy       2004 TCTCAAAAAAAATAAATAAATAAAAT 2028
              |||||||||| ||| ||| |  ||
Db      33894 TCTCAAAAAAAAAAAAAAAAAAGCAT 33870


RESULT 9
ABK83569
ID   ABK83569 standard; cDNA; 122888 BP.
XX
AC   ABK83569;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #140.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 140; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 122888 BP; 28761 A; 33410 C; 31919 G; 28798 T; 0 U; 0 Other;

  Query Match             16.7%;  Score 417.2;  DB 6;  Length 122888;
  Best Local Similarity   61.1%;  Pred. No. 8.7e-76;
  Matches  894;  Conservative    0;  Mismatches  503;  Indels   65;  Gaps   11;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
              ||| |||||  | || || | | |||||  |||||| |||||||||||||  |||  ||||
Db      70248 TTTCTTTTTTATTTTTTTAACACACAGGTCCTTGCTTTGTCACCCAGGCTGGAGTGCAGT 70307

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
              ||||||| || || | || ||| |||||| | |||| |||||||  |||      ||||||
Db      70308 GGTGCAAACACAGCTGACAGCAACCTCAAAC-CCTGGGCTCAAGCAATCCTTCCACCTCA 70366

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAA----------- 772
              || ||||||| ||||||||||||  || ||   ||||| |||  |||||
Db      70367 GCTTCCCAAGAAGCTGGGACTACAGGCATGTGACACCATGCCCAGCTAATTTTTTTTTTT 70426

Qy        773 -----TTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTT 827
                  ||  ||||||||||      |    | |||| |||| |||  |||| ||||
Db      70427 TTTACTTTTAAAAATTTTTTGTAGAAATGGGAGTCTCACTATATTGCTTAGGCTGGTCTC 70486

Qy        828 AAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGG 887
              ||| || |||  | |||||||||||||| ||| ||| |||||||||  |||||||| |||
Db      70487 AAATTCTTGGGCTTAAGCAATCCTCCTGCCTCGGCCTCCCAAAGTTGTGGGATTATAGGT 70546

Qy        888 GTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 947
              |||    ||||     | |  | |   | | ||   |      |   |||| || | |
Db      70547 GTG----ACCACTGCAGCCAGCCTTAATTTCTAATATGGTAAGTATTAATAGCATTTTGG 70602

Qy        948 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007
                | || |         | | || | ||  ||| ||||||||||  | ||||  |
Db      70603 GGATGCTCAATAAT-----TTTCTTCTTCTTTTTTTTTTTTTTTTGAGATGGGGTCTAGC 70657

Qy       1008 TTCATCACCTCAACCTCCG--------------TTCAGCTCACTGCAACCTCTGCCTCCC 1053
                ||||||  ||    |                |||||||||||||||||||||||||
Db      70658 ACTGTCACCTAGACTGGAGTGCAATGGCGCGATTTCAGTTCACTGCAACCTCTGCCTCCT 70717

Qy       1054 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 1113
              |||||||| ||||||| ||||| |||||||||||||||| | |||| ||   | |||
Db      70718 GGGTTCAAGAGATTCTCGTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCACCCGCCA 70777

Qy       1114 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 1173
              |||     |||||||||||||||||||| |||||| ||||||   | |||||| |||||
Db      70778 CCACATCCGGCTAATTTTTTGTATTTTTAGTAGAGATAGGGTTTCGCTATGTTGGCCAGG 70837

Qy       1174 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1233
              |||||||||||||||||||||  |||||| | ||||||||| ||||||||||||||| |
Db      70838 CTGGTCTCAAACTCCTGACCTC--GTGATCCGACTGCCTCAGCCTCCCAAAGTGCTGGGA 70895

Qy       1234 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1293
              ||| ||| ||||| |||    |  ||          | ||  | |||   | ||||
Db      70896 TTACAGGTGTGAGCCACCACACTCGG-----------CAATAATTTTTTAAGGTATAAAA 70944

Qy       1294 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1353
                 |||  ||          | |       |    | |||| |||| ||||       |
Db      70945 ATGTCCAGAGACCAAAAAGGTCAAGAACCACTGCTTTAAATACATAAATGAT-----ATC 70999

Qy       1354 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1413
              |  |||  ||||   ||||    |   |||| |||| | ||   |||       ||  |
Db      71000 AAACTATTTATATTATTTCAAAATTACTTTTAAAAAAATTCAACACAGAGCATAAG-CTC 71058
```

```
Qy       1414 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1473
Db      71059 CATGAAGCAGAGGTTTTTCTTTTTGTTTTATTCATTATGGAAACTAAGTCCCTAGAACAA 71118

Qy       1474 TCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGC 1533
Db      71119 TGCCTGGAACATAGTGGGTGTTCAATAAATACTGATTGAATGAATAAAGAATTATATTTT 71178

Qy       1534 TTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGA 1593
Db      71179 GGGTTTTACCCATTCTGATACGTGGACCTCAAACTCATTCTGCTGTTCAGTATTCTATTA 71238

Qy       1594 AAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTT 1653
Db      71239 TTCCTGAAGATGAAAATATCACAATAACAATTCTATTAATGAACATTTGTTCATAATTTT 71298

Qy       1654 TTTTTAA----TGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1709
Db      71299 TCATGGACATTTTTATACACAGGTAAGAGTTTCTCCAGGGTATATACCTAAGAATGAAAC 71358

Qy       1710 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1769
Db      71359 TGTAGTAAATAAATCTGGCCAGGCACGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA 71418

Qy       1770 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1829
Db      71419 GGCCGAGGCAAGTGGATCAC--AAGGTCAGGGGATCAAGACCATCCTGGCCAACATGGTG 71476

Qy       1830 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1889
Db      71477 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGGGGCGTGTGCCTGTAGT 71536

Qy       1890 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1949
Db      71537 CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGTTTGAACCCAGGAGGTGGAGGTTGCA 71596

Qy       1950 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 2009
Db      71597 GTGAGCCAAGATCGCACCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAC 71656

Qy       2010 AAAAATAAATAAATAAAATAAA 2031
Db      71657 AAACCAAAAAGAAAAAAAAAAA 71678
```

RESULT 10
AAF97850/c
ID AAF97850 standard; DNA; 6405 BP.
XX
AC AAF97850;
XX
DT 31-MAY-2001 (first entry)
XX
DE Human neuroblastoma cell line NB-1 1p36 nucleotide sequence SEQ ID NO:64.
XX
KW Human; chromosome 1; 1p36; neuroblastoma cell line; NB-1; anticancer;
KW tumour suppressor; human 1p36 homozygosity deletion domain; tumour;
KW diagnosis; ds.
XX
OS Homo sapiens.
XX
PN WO200116311-A1.
XX
PD 08-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-JP005930.
XX
PR 31-AUG-1999; 99JP-00245962.
PR 09-MAY-2000; 2000JP-00136266.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
PI Nakagawara A;
XX
DR WPI; 2001-226686/23.
XX
PT Human 1p36 homozygosity deletion domain from the 36-position of first
PT chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT in gene diagnosis of tumors as well as in developing anti-cancer drugs.
XX
PS Example 8; Page 93-95; 226pp; Japanese.
XX
CC The present invention describes a homozygosity deletion domain co-
CC existing in the 36-position of the first chromosome short arm (1p36) in
CC human neuroblastoma. Also described are base sequences from the 1p36
CC position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC which are tumour suppressor genes in human neuroblastoma. The genes are
CC tumour suppressor genes, base sequence data of which are applicable as
CC tumour markers and reagents in studying mechanism of tumour body
CC formation, and gene diagnosis of tumours as well as in developing anti-
CC cancer drugs. AAF97787 to AAF97829 represent PCR primers used in the
CC exemplification of the present invention, and AAF97830 to AAF97874
CC represent sequences given in the exemplification of the present invention
XX
SQ Sequence 6405 BP; 1650 A; 1449 C; 1383 G; 1923 T; 0 U; 0 Other;

Query Match 16.7%; Score 416.2; DB 5; Length 6405;
Best Local Similarity 59.9%; Pred. No. 7.5e-76;
Matches 856; Conservative 1; Mismatches 504; Indels 68; Gaps 7;

```
Qy        603 GTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAG 662
Db       3636 GTCTTTATTTTTTATTTTTTTGAGGTGAGTTTCACTCTTGTTGCCCAGGCTGGAGTGCAA 3577

Qy        663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 722
Db       3576 TGGCACGATCTCAGCTTACAACAACCTCTGCCTCCTGAGTTCAAGTGATTCTCCTGCCTC 3517

Qy        723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db       3516 AGCCTCCCGAGTAGCTGGGATTACAGGCATGCACCACCACACTCAGCTGATTTTTTTGTA 3457

Qy        783 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db       3456 TTTTTAGTAGAGA-TGGGGTTCTCCATGTTGGTCAGGCTAGTCTCAAACTCCCGACCTCA 3398

Qy        843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db       3397 GGTGATCCGCCTGCCTCGGCCTCCCGAAGTGCTGGGATTACAGGCGTGAGCCATAAGGCC 3338

Qy        903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db       3337 TGGCCTATTTTTTACT---------------------------------------------- 3322

Qy        963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db       3321 -TTTAAGATGGAGTTTTGCTCTTGTTGCCCAAGCTGGAGTGCAATGGCATGATC------ 3269

Qy       1023 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 1082
Db       3268 -----TCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGTGATTCTCTTGCCTCAGTCT 3214

Qy       1083 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTA 1142
Db       3213 CCCAAGTAGCTGGGATTACAGGCATGCACCACCATGCCCGGCTAA--ATTTGTATTTTTA 3156

Qy       1143 GCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGAT 1202
Db       3155 GTAGAGATGGGGTTTCTCCATGTTGGTTAGGCTGGTCTCAAACTCCCAACCTC-AGTGAT 3097

Qy       1203 CTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
Db       3096 GCACCAGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGCCG 3037
```

```
Qy       1263 ATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAG 1322
Db       3036 TACGTATCTTTATATCAGGATGTTTATAACAATTTTCTCTTAGAACATCTTTAGAATTTA 2977

Qy       1323 TAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTT 1382
Db       2976 ATACTAGGAAAAAAGGTTT-------CTCTCGACAAACTATTAGGCTTTCTTTTACCAGA 2924

Qy       1383 TTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACT 1442
Db       2923 GTGAAAATTGTACAATTTATCTTTCCCTTTTTGCATTAAATTCCATGTTCCAAATTAAAT 2864

Qy       1443 TATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCAT 1502
Db       2863 TACAATAACCACATTAGCCGAGGCAACCTATACATTTAGTTTTTCTTGTTACATAAATTT 2804

Qy       1503 TGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCA 1562
Db       2803 TTACTTATTTATATTTTAATAGCTTCACTGTATATATTTGTGGTAAGCTGCTCAAATCCT 2744

Qy       1563 CTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATA 1622
Db       2743 TTCTAGAAGTAGTAACTGGTATAAATTATAAATAAATAAAAACTTCCTGTAATAGCCTAT 2684

Qy       1623 CACACACACACACACACTC-ACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA 1681
Db       2683 TCTGCTATTTAACACTGTAAATATGTTGTGTAGTCAAAGTTCTAGTTCCTAAAAACAAAT 2624

Qy       1682 ACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGAC 1741
Db       2623 AAAACCAAACGACACAGCAAGCTTTTAAAAGTTACATAAAGATCGGCTGGGCATCGTGGC 2564

Qy       1742 TCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAA 1801
Db       2563 TCACGCGTATAATCCCAACACTTTGGGAGGCCAAGGCGGGTGGATCACTTGAGGTCAGGA 2504

Qy       1802 GTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTA 1861
Db       2503 GTTCAAGACCAGCCTGGCCAACATGATGAAACGCTGTCTCTACTAAAAATACAAAAATTA 2444

Qy       1862 GCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAAT 1921
Db       2443 GCTGGGTGTGATGGCACATGACTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT 2384

Qy       1922 TGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCC 1981
Db       2383 CACTTGAACCCGGGAGATGAAGGTTGCAGTGAGCCAAGGTCACACCACTGCACTCTAGCC 2324

Qy       1982 TGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db       2323 TGGATGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAGTTGCACAAA 2275
```

```
RESULT 11
ADP13332/c
ID   ADP13332 standard; DNA; 126990 BP.
XX
AC   ADP13332;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Renal cell carcinoma differentially expressed gene #68.
XX
KW   ds; diagnosis; non-blood disease; solid tumor; gene expression;
KW   peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW   head/neck cancer; differential expression.
XX
OS   Homo sapiens.
XX
PN   WO2004048933-A2.
XX
PD   10-JUN-2004.
XX
PF   21-NOV-2003; 2003WO-US037481.
XX
PR   21-NOV-2002; 2002US-0427982P.
PR   03-APR-2003; 2003US-0459782P.
XX
PA   (AMHP ) WYETH.
PA   (TWIN/) TWINE N C.
PA   (BURC/) BURCZYNSKI M E.
PA   (TREP/) TREPICCHIO W L.
PA   (DORN/) DORNER A.
PA   (STOV/) STOVER J A.
PA   (SLON/) SLONI D K.
XX
PI   Twine NC,  Burczynski ME,  Trepicchio WL,  Dorner A,  Stover JA;
PI   Sloni DK;
XX
DR   WPI; 2004-460799/43.
XX
PT   Diagnosing non-blood disease such as solid tumor, involves comparing
PT   differential expression profile of specific genes in peripheral blood
PT   sample of subject with reference expression profile of specific genes.
XX
PS   Disclosure; SEQ ID NO 68; 350pp; English.
XX
CC   The invention relate to a method of diagnosing (M1) non-blood disease
CC   such as solid tumor by providing peripheral blood sample of human having
CC   non-blood disease, and comparing an expression profile of specific genes
CC   in the peripheral blood sample to reference expression profile of the
CC   genes, where each of the genes is differentially expressed in peripheral
CC   blood mononuclear cells (PBMCs) of patients having the disease as
CC   compared to PBMCs of normal humans. The method is useful for diagnosing
CC   non-blood disease such as solid tumor. The solid tumor is chosen from
CC   renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC   peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC   sample is a whole blood sample (claimed). (M1) is useful for identifying
CC   genes that are differentially expressed in peripheral blood samples
CC   isolated at different stages of progression, development or treatment of
CC   RCC and/or other solid tumors. This sequence corresponds to a gene that
CC   is differentially expressed and detected by the method of the invention.
CC   (Note: this sequence is not given as part of the printed specification
CC   but was obtained from WIPO in electronic format at
CC   ftp.wipo./pub/published_pct_sequences).
XX
SQ   Sequence 126990 BP; 36683 A; 27389 C; 27065 G; 35853 T; 0 U; 0 Other;
```

```
  Query Match             16.6%;  Score 416;  DB 12;  Length 126990;
  Best Local Similarity   60.2%;  Pred. No. 1.5e-75;
  Matches  859;  Conservative    1;  Mismatches  536;  Indels   32;  Gaps    9;

Qy        611 TTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAA 670
Db      27332 TTTCTCTAATGCTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGA 27273

Qy        671 CCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCC 730
Db      27272 TCTCA-----------CCTCCACCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCT-CC 27225

Qy        731 AAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGT 790
Db      27224 TGGTAGCTGGGAATACAGGCGTGCACCACCATGCCCGGCTAATTTTTATA--TTTTCAGT 27167

Qy        791 AGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCC 850
Db      27166 AGAGTCGCAGTTTCACCATGTTTCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCC 27107

Qy        851 TCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACT 910
Db      27106 ACCCGCCGCAGCCTCCCAACGTGCTGGGATAATAGGGGTGAGCCACTGCACCTGGCCTTT 27047

Qy        911 TATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGC 970
Db      27046 CTGTACTCTTTATTCTATTTCACTGGTCTGTTTCTTTTGGTTTCAGTATCATAGTACTTT 26987
```

```
Qy       971 CTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA 1030
Db     26986 TGTTTCTGTAGCTTAAAATAT--ATATATTTTAAGACTGGAATGCAGTGGCATGATCTCA 26929

Qy      1031 GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTA 1090
Db     26928 GCTCACTGCAACCTCCACCTCCCAGGCTCAAATTATTCTTGTGACTCAGCCTCCTGAGTA 26869

Qy      1091 GCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGAT 1150
Db     26868 GCTGGGATTACAGGCAGGCACCACCATGCCCCACTAA-TTTTTGTATTTTCAGTAGAGTG 26810

Qy      1151 GGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGC 1210
Db     26809 GAAGTTTCACCATGTTTCCTAGGCTGATCTTAAACTCCTGACCTCGAGTGATCCGCCCGC 26750

Qy      1211 CTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGT 1270
Db     26749 GTCAGCCTCCCAAAGTGCTGGGATAACAGGCATAAGCTGCTGCGTGCAGCCTTTCTGTAT 26690

Qy      1271 CTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCT 1330
Db     26689 TCTTTATTCTGTTTCATTGATCTGTTCCTTTTGGTTTCAGTATCATACTATTTTTGTTTT 26630

Qy      1331 CAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAAC 1390
Db     26629 GTTTCTATAGCTTAAAAATATGTATTTT---TTTTGAGATAGGGTCTCACTGTGTCACCC 26573

Qy      1391 AATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCT 1450
Db     26572 AGGCTGGAGTGTACTGGTGTGATCATGGCTCACTGCATCCTCAACCTCCTATACTCAA-T 26514

Qy      1451 TGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTA 1510
Db     26513 TGATCCTCCCACCTCAGCCTCCCAAGTAGCTGCACCACCATGTTTGGCTAATTTTTAAAT 26454

Qy      1511 TTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAA 1570
Db     26453 TTTTTGTAGAAACAGAGTCTTGCCATGTCACCCAGGCTGGTCTTGAACTCCTGGGCTGAA 26394

Qy      1571 GGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACA 1630
Db     26393 G---AGATCCGCTGACCTCATCCTCCCAAAGTGCTGGAATTACAGGTGTGCACCACCTTG 26337

Qy      1631 CACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA--------A 1682
Db     26336 CCCAGCCAAAAATATGTTTTTGGCAGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGG 26277

Qy      1683 CCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACT 1742
Db     26276 CCAACATGGCAAAACCTTGTCTCTACTAAAGAATACAAAAATTATCTGGGCGCAGTGGCT 26217

Qy      1743 CAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAG 1802
Db     26216 CATGTCTGTAAATCCAGTACTCTGGAAGGCCAAGGCAGGTGGATCACCTGAGGTCAGGAG 26157

Qy      1803 TTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAG 1862
Db     26156 TTTAAGACCAGCCTGGCCAACCTGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAG 26097

Qy      1863 CTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATT 1922
Db     26096 CTGGGCGTGGTGACCGGTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 26037

Qy      1923 GCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCT 1982
Db     26036 ACTTGAACCTGGGAGGTAGAGGTTGCAGTGAGCCGAGATCATGCCATTGCACTCTAGCCT 25977

Qy      1983 GGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db     25976 GGGTGACAGAGCGAGATTCTGTCTCAAAAAAGAAAAAAAGAAAAAAA 25929
```

```
RESULT 12
ACN44170/c
ID   ACN44170 standard; DNA; 196686 BP.
XX
AC   ACN44170;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG39530.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 484; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 196686 BP; 53978 A; 42758 C; 43862 G; 55372 T; 0 U; 716 Other;

  Query Match             16.6%;  Score 414.2;  DB 11;  Length 196686;
  Best Local Similarity   60.7%;  Pred. No. 4e-75;
  Matches  869;  Conservative    1;  Mismatches  464;  Indels   97;  Gaps    8;

Qy       605 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 664
Db     98993 TTCATTCATTTATTCATTCGAGATGGGGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAATG 98934

Qy       665 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 724
Db     98933 GTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGGATTCAAGTGATTCTCCTGCCTCAG 98874

Qy       725 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 784
Db     98873 CCTCCTGAGTAACTGGGATTACAGACACCTGCCACCACACCCGGCTAATTTTTGTA--TT 98816

Qy       785 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 844
Db     98815 TTTAGTAAAGATGGAGTTTCACCATGTTGGCCAGCCTGGTCTCAAACTCCTGACCTCAAA 98756

Qy       845 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 904
Db     98755 GGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCA 98696
```

```
Qy       905 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 964
Db     98695 GCCATGTACATTTATAATTGTACTATATCTCCTTTGGT--------CACCATCTTGTCTT 98644

Qy       965 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 1024
Db     98643 TTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCGCCCAGGCTGG 98584

Qy      1025 CGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTC 1070
Db     98583 AGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCTGGCTCACGCCATTCTC 98524

Qy      1071 CTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTT 1130
Db     98523 CTGCCTCAGCCTCCAGAGTAGCTGGGACTACATGTGCCCACCACCACGCCCGGAGAATTT 98464

Qy      1131 TTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTG 1190
Db     98463 TTTGTATTTTTAATGGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTG 98404

Qy      1191 ACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCAC 1250
Db     98403 ACCTC--GTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGAGTGAGCCAC 98346

Qy      1251 TGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTG 1310
Db     98345 TGGGCCCGGCCACCACCTCGTCTTCTGATTGCTTTTTTTTTTCACTTCTGGAGAG----- 98291

Qy      1311 TCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTT 1370
Db     98290 ----------------------------------------------------AGGAAATC 98283

Qy      1371 TCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTC 1430
Db     98282 TATATGTTCTTCACCATATCTCTAATACCTTCTTTACCTTCTCTCTAATATCTAATACTA 98223

Qy      1431 TGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTA 1490
Db     98222 TATATACATACATATAGTGTGTATATATATATGCATGTATATATAGTATTATATGTATAT 98163

Qy      1491 CTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACAT 1550
Db     98162 ATATAATAACCAGGTTCTTATTTATGTTATCCATTGCTTGAATTAACAATGTTTTTTCA- 98104

Qy      1551 CCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTC 1610
Db     98103 -------AAACCTATAAAAAGACCTTAAACAAGGTAAGGATGAGAGGAACCTTGAGTTGC 98051

Qy      1611 CCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAAC 1670
Db     98050 CCAAGATAAAATTTTGTCTCAAAGTCTCTCACTTTGTTTCATGTGTTTAATCTCTCAATA 97991

Qy      1671 TAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAA----CTCA 1726
Db     97990 TATTTAATCTATTGGCAATTACAATAATTTTTGTTTTGGTCACAAAAAATAAGCTCCTGG 97931

Qy      1727 GTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGAT 1786
Db     97930 GCTGGACACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGAGGGTGGAT 97871

Qy      1787 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1846
Db     97870 CAC--CAGGTCAGGAGATGGAGACCATCCTGGCCAACATGGTAAAACTCTGTCTCTACTA 97813

Qy      1847 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1906
Db     97812 CAAATACAAAAATTAGCTGGGCGTGGTGGCATGCGCCTGTAGTCCCAGCTACTCGGGAGG 97753

Qy      1907 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1966
Db     97752 CTGAGGCAGGAGAACTGCTTGAACCTGGGAGGCAGACGTTGCAGCGAGCTGAGATCACGC 97693
```

```
Qy      1967 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAA 2017
Db     97692 CACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCTATCTCAAAAAATAAA 97642
```

RESULT 13
ADP68568/c
ID ADP68568 standard; cDNA; 86000 BP.
XX
AC ADP68568;
XX
DT 09-SEP-2004 (first entry)
XX
DE Human PPAR-alpha cDNA.
XX
KW cytostatic; gene therapy; PPAR-alpha;
KW peroxisome proliferator-activated receptor-alpha; PPAR-alpha modulator;
KW PPAR-alpha associated disorder; hyperproliferative disorder; human; gene;
KW ss.
XX
OS Homo sapiens.
XX
PN US2004115637-A1.
XX
PD 17-JUN-2004.
XX
PF 11-DEC-2002; 2002US-00317500.
XX
PR 11-DEC-2002; 2002US-00317500.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Mckay R, Dobie KW;
XX
DR WPI; 2004-449378/42.
XX
PT New oligonucleotide compound that inhibits expression of PPAR-alpha,
PT useful for preparing a composition for treating hyperproliferative
PT disorders, e.g. cancer.
XX
PS Claim 1; SEQ ID NO 4; 121pp; English.
XX
CC The invention describes a compound, having a sequence comprising 8-80 bp
CC targeted to a nucleic acid encoding PPAR-alpha (peroxisome proliferator-
CC activated receptor-alpha), that specifically hybridises with the nucleic
CC acid encoding PPAR-alpha comprising 86001-bp sequence and inhibits
CC expression of PPAR-alpha. Also described are: a method of inhibiting the
CC expression of PPAR-alpha in cells or tissues; a method of screening for a
CC modulator of PPAR-alpha; a diagnostic method for identifying a disease
CC state; a kit or assay device comprising the compound; and a method of
CC treating an animal having a disease or condition associated with PPAR-
CC alpha. The oligonucleotide compound is useful for preparing a composition
CC for treating hyperproliferative disorder e.g. cancer. This sequence
CC represents a human peroxisome proliferator-activated receptor-alpha (PPAR
CC -alpha) cDNA.
XX
SQ Sequence 86000 BP; 22822 A; 19623 C; 20377 G; 23178 T; 0 U; 0 Other;

  Query Match   16.5%; Score 413.6; DB 12; Length 86000;
  Best Local Similarity 61.0%; Pred. No. 4.5e-75;
  Matches 878; Conservative   1; Mismatches 490; Indels   71; Gaps   10;

```
Qy       605 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 664
Db     69228 TTTTTTTTTCTTTTTTTTTGAGACGGAGTCGCACTCTGTCGCCCAGGCTGGAGTGCAGTG 69169

Qy       665 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 724
Db     69168 GTGCGATCTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCTATTCTCCTGTCTCAT 69109

Qy       725 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 784
Db     69108 CCTCCCAAGTAGCTGGGATTACAGGTGTGCACTACCATGCCCAGCTAATTTTTGTA--TT 69051
```

```
Qy       785 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 844
Db     69050 TTTATTAGAGACGGGGTTTTGCCATGTTCGCCAGGCTGGTCTCAAACTCCTGACCTCAAG 68991

Qy       845 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 904
Db     68990 TGATCCACCCGCCTCGGCCT-CCAAAGTGCTGGGATTACAGGTGTGAACCACCGCACCTG 68932

Qy       905 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 964
Db     68931 GGAGACTTTTATTTTCTTTCTTTCTTTCTTTTTCTTTTTTTTTTTTTTGTGAGACGGAAT 68872

Qy       965 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 1024
Db     68871 CTCCCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCAATC---------------------- 68833

Qy      1025 CGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 1084
Db     68832 ---TCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTTCTGCCTCAGCCTCC 68776

Qy      1085 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 1144
Db     68775 CGAGTAGCTGGGACTACAGGCATGTGCCACCACACCCAGCTAA-TTTTTGCATTCTTAGT 68717

Qy      1145 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1204
Db     68716 AGAGACGGGGTTTCACCATATTGGCCAGGCTGGTCTCAAACTCCTGACCTGAGGTGATCC 68657

Qy      1205 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGAT 1264
Db     68656 ACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAACCACCGCGCCTGGGCGA- 68598

Qy      1265 TACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTA 1324
Db     68597 ----CTTTTATTTTCTTAGGCCACTCTGATTACAAAATGACTTGGGTGTGAAATGTTGAA 68542

Qy      1325 GGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTT 1384
Db     68541 GCACATTCATCCATAGCTTGATCTGTTCCCTTGGAGCAAGGAAGGGGAACTGA------- 68489

Qy      1385 TAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTA 1444
Db     68488 -GGAACGAACTGGGAAAATGTGCAG------GGCCACCTTACCTACCGTTGTGTGACATC 68436

Qy      1445 TCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTG 1504
Db     68435 CCGAC-AGAAAGGCACTTGTGAAATCGACAATACTGGCATTTGTTTCTGTTCTTTTTCTG 68377

Qy      1505 TTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACT 1564
Db     68376 ----GATCTTGCAGCTGCGGTCGCACTTGTCATACACCAGCTTGAGTCGAATCGTTCGCC 68321

Qy      1565 TTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACA 1624
Db     68320 GAAAGAAGCCCTGTGAGGGTAATACACAGTGGATTTAAAGACCTGGGCCACTGAGAAACC 68261

Qy      1625 CACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACC 1684
Db     68260 AGCGCACTTGCTTTACTCACCTTTTTTGTTGCTCATTTATTGTTTCTTTTTAAATGAACA 68201

Qy      1685 TGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCA 1744
Db     68200 AAATTTAGAGAAAAGTTATTC------------------CAGGCCAGGCATGGTGATCA 68160

Qy      1745 AGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTT 1804
Db     68159 TGCCTGTAATCCCAGCACTTTAGGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTT 68100

Qy      1805 CGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCT 1864
Db     68099 CGAGACCAGCCTAGCCAACATGGTGCAACCCTGTCTCTACTAAAAATACAAAAATTAGCC 68040
```

```
Qy      1865 GGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGC 1924
Db     68039 GGGCATGGTGATGTAAGTCTGTAGTCCCAGCTACTCAGGAGGCTGAGGTAGGACAATCAC 67980

Qy      1925 TTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGG 1984
Db     67979 TTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGC 67920

Qy      1985 GCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAA 2044
Db     67919 GCGACAGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGAAA 67860
```

RESULT 14
ADQ97523/c
ID ADQ97523 standard; DNA; 215974 BP.
XX
AC ADQ97523;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD09-008, SEQ ID 500.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 500; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;

Query Match 16.5%; Score 413.6; DB 12; Length 215974;
Best Local Similarity 59.6%; Pred. No. 5.4e-75;
Matches 847; Conservative 0; Mismatches 529; Indels 46; Gaps 7;

```
Qy       610 TTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCA 669
Db      4846 TGTCTTTTGTTTTTGAGACAGGGTCTCGCTCTGTTGCCCAGGCTGGAGTACAGTGGCACA 4787

Qy       670 ACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCC 729
Db      4786 ATCATGGCTCATTGCAGCCTCAAACTCCTGGGCTCAAATGAGCCTCCAACCTCAGCCTCC 4727

Qy       730 CAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTG 789
Db      4726 TAAGTAGCTGGAACTACAGGCATACATCACCACTCCTAGCTAATTTTTGTATTTTTT--- 4670

Qy       790 TAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATC 849
```

```
                    |||| |||  |||| |||||||| ||||  || |||| |  ||||||||||
Db       4669 ----------GTCTCACTGTGTTGCCCAGGCTGGTCTCCAATTCCTAGGCTCAAGCAATC 4620

Qy        850 CTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTAC 909
              |   |    || |||||||||| ||||||||||   |||||||  |     | |||
Db       4619 ACCTGCCTCAAGCCTCCCAAAGTACTGGGATTACAGACATGAGCCACGACACCCAGCTTG 4560

Qy        910 TTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTG 969
              | |  ||    |  | ||   ||||  | |    |              ||
Db       4559 AAAGTGTCTTTTTATTATTATTATTAATATTTTTTTTTTTTTGAGATGAAG--------- 4509

Qy        970 CCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTC 1029
                       | ||  | |   | |   ||   | | ||    | ||| ||| | ||
Db       4508 ------------TCTCGCTCTGTTGCCCAGGTTGGAGTACAGTGGCGCAATCTCGGCTC 4462

Qy       1030 AGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGT 1089
              | | |   |||||||||||||||| |||||  | ||||||||||||| ||||||| ||||
Db       4461 ACCGC---GCAACCTCTGCCTCCCAGGTTCCGGGGATTCTCCTGCCTCAGCCTCCCGAGT 4405

Qy       1090 AGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGA 1149
              |||||| ||||||||| ||| |||||| ||   ||||| |||||||| |||||| |||||
Db       4404 AGCTGGGATTACAAGCATGCGCCACCACGCCCAGCTAA-TTTTTGTAGTTTTAGTAGAGA 4346

Qy       1150 TGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTG 1209
              | || ||| | ||||||| ||||||||||||  ||||||||||||| ||||||  ||
Db       4345 TAGGCTTTCATCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCACCCA 4286

Qy       1210 CCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTG 1269
              |||||  |||||||||||||| |||| |||| |||    |   | | | |    |
Db       4285 TCTCAGCTTCCCAAAGTGCTGGGATTACAGGCATGACCACCGCACCTGGCCGAAAGTATC 4226

Qy       1270 TCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGC 1329
              | | |   | | ||| | |   |  | | || |   |||| |||  |  |||||||  |
Db       4225 TTTTTAAAGTCTCTTGATCT---CTAACACCTCGTATAGTGGCTGGTACTTAGTAGGCAC 4169

Qy       1330 TCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAA 1389
              |   |||   ||||  ||||| | |   ||   | |      |  |  |   | |
Db       4168 T---TAACACATTGGAGAATGAAGATATTAACCACATGGCATTATCATTACTCTCAGCTT 4112

Qy       1390 CAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACC 1449
              |||             ||     |  |   | | |||   |     | ||   |  |
Db       4111 AAATGGCTTGCCTCAAGCCACGCAGCTGGTAGGTGGCAGAGTTGGAATGCAAACACAAAT 4052

Qy       1450 TTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACT 1509
              |||||          || ||  | |   ||   |     |  |   |     ||||
Db       4051 TTGTTTGATTCCACAAATCTTTGCCCTTTCCACAAACAAGGCCCTACATGTCCCCTTACA 3992

Qy       1510 ATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGA 1569
              |||      || | | |||   |  |      |  || ||||  ||     ||  |
Db       3991 TATTTTTCAGGTGTACATTATCTTTCCAGCATTGTGAAC-TCCTAGTGAAAGCCTAGATC 3933

Qy       1570 AGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACAC 1629
              | ||    |   | || | |      |      |     |  | |    ||  |
Db       3932 TCACTAATAGCAATGACCTCTTCTAGCTGTTATTTATTGAGTGCTTATGATGTGCCAGGC 3873

Qy       1630 ACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCAT 1689
              |      |||      | |  |     ||     |  ||||       ||
Db       3872 ATTGCATTTAGCACTTCACACATATGATCTCATTTAATTTTCACAACAATCCCATGAGAT 3813

Qy       1690 TAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCT 1749
              | |       | |  |     || |  | |||  |  |||    |||  ||| |||| ||||
Db       3812 TGGCACTATTATTATGGCCTTTAACAGAAAGGAAACAGGCCAGCATGGTGGCTCATGCCT 3753

Qy       1750 GTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGA 1809
              |||  |  || ||||||| || |  |||||||| |||||| |||||| || ||||| |||
Db       3752 ATAATCCTAGCACTTTGGGAGGCTGAGGTGGGTTGATCACCTGAGGTCAGGAGTTCAAGA 3693

Qy       1810 CCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTG 1869
              ||| |||||  ||| || |||||||| ||||||||||||||||||||||||||  |||
Db       3692 CCAACCTGGCCAACATAGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCA 3633

Qy       1870 TAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAA 1929
              | |||  || |||||||| |||||||||||||||||||||| || ||||||||||||||
```

```
Db       3632 TGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAA 3573

Qy       1930 CCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGAC 1989
              ||  |||||| |||||||||||||||| |||||  |||   |||||||||||||||| |||
Db       3572 CCCAGGAGGCGGAGGTTGCAGTGAGCAGAGATTGCACTGTTGCACTCCAGCCTGGGAGAC 3513

Qy       1990 ACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 2031
              | ||| ||||||  |||||||| | | | | |  ||| | |
Db       3512 AGAGCAAGACTCCGTCTCAAAAAGAAACAGAGAGGAAACAGA 3471


RESULT 15
AAL42369/c
ID   AAL42369 standard; DNA; 13670 BP.
XX
AC   AAL42369;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human Guanine nucleotide binding protein gamma 7 (GNG7) gene sequence.
XX
KW   Human; gene; ds; guanine nucleotide binding protein gamma 7; G protein;
KW   GNG7; novel polymorphic site; drug screening; gene therapy;
KW   GNG7-related disease; pancreatic cancer; GNG7 haplotyping;
KW   GNG7 genotyping.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   allele          replace(3900, T)
FT                   /*tag=  a
FT                   /note= "Single nucleotide polymorphism"
FT   allele          replace(3975, G)
FT                   /*tag=  b
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4005. .9671
FT                   /*tag=  c
FT                   /product= "Guanine nucleotide binding protein gamma 7"
FT   exon            4005. .4085
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4086. .9545
FT                   /*tag=  e
FT                   /number= 1
FT   exon            9546. .9671
FT                   /*tag=  f
FT                   /number= 2
FT   allele          replace(9958, A)
FT                   /*tag=  g
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200218647-A1.
XX
PD   07-MAR-2002.
XX
PF   23-AUG-2001; 2001WO-US026279.
XX
PR   25-AUG-2000; 2000US-0228234P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Finkel K,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-315548/35.
DR   P-PSDB; AAO14753.
XX
PT   Novel genetic variants of guanine nucleotide binding protein, gamma 7
PT   gene useful in studying expression and function of the protein, and for
PT   screening drugs to treat diseases e.g. pancreatic cancer.
XX
PS   Claim 1; Fig 1; 62pp; English.
XX
CC   The invention comprises the nucleotide and amino acid sequences of the
```

CC human guanine nucleotide binding protein (G protein) gamma 7 (GNG7). The
CC invention specifically relates to the discovery of three novel
CC polymorphic sites in the GNG7 gene. The GNG7 nucleotide and protein
CC sequences are useful for screening for drugs which target GNG7 and may be
CC used to treat (gene therapy) GNG7-related diseases (e.g. pancreatic
CC cancer). The GNG7 nucleotide sequence can be used for haplotyping GNG7,
CC genotyping GNG7 and predicting the haplotype pair for GNG7 in an
CC individual. The present DNA sequence represents the human GNG7 gene
CC sequence
XX
SQ Sequence 13670 BP; 3047 A; 3946 C; 3680 G; 2997 T; 0 U; 0 Other;

Query Match 16.5%; Score 413.4; DB 6; Length 13670;
Best Local Similarity 60.6%; Pred. No. 3.3e-75;
Matches 876; Conservative 1; Mismatches 512; Indels 56; Gaps 10;

```
Qy       604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCAT-GAGCACAG 662
Db      5900 TTTGTTTTTGTTTTTTTTCTGAGATGGAGTCTCACCCTGTCGCTTGGGGTTGGAGGGCAG 5841

Qy       663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 722
Db      5840 TGGCGCAATCTCAGCTCACTGCAACCTCTGCCTCTCGGGTTCAAGCTATTCTCCTGACTC 5781

Qy       723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db      5780 AGTCTCCCAAGTAGCTGGGACTATAGGAGCCTGCCACCACGCCTGGCTAATTTTTTTGTA 5721

Qy       783 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db      5720 ATTTTAGTAGAGATGGGGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTGAC--CT 5663

Qy       843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db      5662 TGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGCGATTCCAGGCGTGAGTCACCGTGCC 5603

Qy       903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db      5602 TGGCCAGTTAAATGTTTTTTTAGATGTTGCTGTGAAGGTTGTTTTTTTC---------- 5554

Qy       963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db      5553 -TTGTTTCTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTCTCGCTCTGTTCCCCAGGCT 5495

Qy      1023 TCCGT----------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 1068
Db      5494 GGAGTGCAATGGTGCGATCTCGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTC 5435

Qy      1069 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA- 1127
Db      5434 TCCTGCCTCAGCCTCCCAAATAGCTGGGATTACAGGCATGCACCACCATACCCGGGTAAT 5375

Qy      1128 ---TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAA 1184
Db      5374 CTTTTTTTTGTATTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA 5315

Qy      1185 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTG 1244
Db      5314 CCCCTGACCTCAGGTGATCTGCCCACCTCAGCTTCCCAAAGTGCTGGAATTATAAGTGTG 5255

Qy      1245 AGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGA 1304
Db      5254 AGTCACCGCGCTCAGCCTGTT-TTTTATTTTGTAGAGATGGGGTCTTGCTATGTTGTTGC 5196

Qy      1305 GCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATAT 1364
Db      5195 CCAGGCTGGAATTGAACGCCTGACCTCAAGTGATCCTCCTGCCTCGGCCTCCCAATGTGC 5136

Qy      1365 AAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTT 1424
Db      5135 TGGGATT----------------ACAGCTGTGAGCCACTGTGCCTCACATGCAAAATTTT 5092

Qy      1425 GCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1484

Db      5091 TTTTTTTTTTTTGATGTGATTCACATT----TAAATCCGTAGACTTCAAGCACAGCAGGT 5036

Qy      1485 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1544
Db      5035 GACCGTCTGTAATGCAGATGGGCCTCATCTAATCAGTTGAAGGGCTTAAGAGAAAAAGCC 4976

Qy      1545 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1604
Db      4975 TGGGGTCCTTCCAGGAGGAGGGAACTCTGCCTCCAGGTGGCCTTGGGGTTCCAGCTGCAG 4916

Qy      1605 CCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTT 1664
Db      4915 CACCAACTCTTCCCTGGGTCTCCAGCCTGCTGGCCTACACTGCAGATTTCAGAGCTTGCC 4856

Qy      1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724
Db      4855 TTAAAATAAATCTCGACAA-GGCGCAGCCTCAAGTGTGATCCTTGTAAAAAACTTCAAAC 4797

Qy      1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784
Db      4796 GGGCCAGGCATGGTGGCTCACGCCTGGAATCCCAGGACTTTGGGAGGCCGAGGCAGGCGG 4737

Qy      1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844
Db      4736 ATCACCTGAAGTTAGGAGTTCAAGACCAGCCTGGGCAACATGGCGAAACGCCGTCTCTAC 4677

Qy      1845 TAAAAATAC-AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGG 1903
Db      4676 TAAAAATACAAAAAATTAGCCAGGCATGGTGGTGCATGCCTGTAATCCCAGCTATTTGGG 4617

Qy      1904 AGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCC 1963
Db      4616 AGGTTGAGGCAGAAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCG 4557

Qy      1964 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 2023
Db      4556 CGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAACAAACAAAACAAA 4497

Qy      2024 AAAAT 2028
Db      4496 CAAAT 4492
```

Search completed: February 9, 2006, 18:54:29
Job time : 1619.05 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 15:02:28 ; Search time 9346.29 Seconds
(without alignments)
12519.885 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 2499.4
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	505.4	20.2	507	5	BU785040	BU785040 in43g09.x
	2	344.2	13.8	4087	4	BC024593	BC024593 Homo sapi
	3	341	13.6	3990	4	HSM803026	AL713681 Homo sapi
c	4	337.2	13.5	15970	9	AQ839852	AQ839852 260L13-C5
c	5	315.8	12.6	897	6	CD557847	CD557847 AGENCOURT
	6	311	12.4	3095	4	CR858326	CR858326 Pongo pyg
c	7	307.2	12.3	5797	4	CR749233	CR749233 Homo sapi
c	8	301.2	12.1	1605	4	CR860521	CR860521 Pongo pyg
	9	301.2	12.1	1863	4	CR599842	CR599842 full-leng
	10	300.6	12.0	946	5	BQ958903	BQ958903 AGENCOURT
c	11	298	11.9	918	5	BQ706343	BQ706343 AGENCOURT
	12	293.6	11.7	736	8	CX780936	CX780936 HESC3_2_B
	13	290.6	11.6	3092	4	CR613629	CR613629 full-leng
c	14	288.2	11.5	855	5	BQ681302	BQ681302 AGENCOURT
c	15	287.6	11.5	769	5	BQ710315	BQ710315 AGENCOURT
c	16	285.2	11.4	4088	4	HSM808459	BX648311 Homo sapi
	17	283	11.3	3165	4	CR859576	CR859576 Pongo pyg
c	18	281.6	11.3	800	1	AU120942	AU120942 AU120942
	19	281.4	11.3	3259	4	CR860168	CR860168 Pongo pyg
	20	280.4	11.2	672	6	CA431692	CA431692 UI-H-DF0-
	21	280	11.2	3474	4	HSM800886	AL110229 Homo sapi
	22	278.8	11.2	658	9	AQ393450	AQ393450 CITBI-E1-
c	23	277.6	11.1	877	9	AQ739838	AQ739838 HS_5505_A
	24	277.2	11.1	1875	4	BC009270	BC009270 Homo sapi
c	25	276.6	11.1	3990	4	HSM803026	AL713681 Homo sapi
	26	276.4	11.1	4828	4	HSM802759	AL390128 Homo sapi
c	27	276.2	11.1	881	5	BU521286	BU521286 AGENCOURT
	28	274.8	11.0	3552	4	CR860263	CR860263 Pongo pyg
	29	273.8	11.0	736	6	CA427039	CA427039 UI-H-DF0-
c	30	273.4	10.9	629	5	BX509360	BX509360 DKFZp686J
c	31	273.4	10.9	2230	4	CR859082	CR859082 Pongo pyg
	32	272.4	10.9	652	6	CA427045	CA427045 UI-H-DF0-
	33	272.4	10.9	666	6	CA431783	CA431783 UI-H-DF0-
c	34	272.4	10.9	815	8	CX785622	CX785622 HESC3_49_
c	35	272.2	10.9	922	5	BU501973	BU501973 AGENCOURT
c	36	272	10.9	617	3	BI861844	BI861844 603388872
	37	271.8	10.9	644	7	CN480313	CN480313 UI-H-EU0-
	38	271.6	10.9	1042	5	BX377759	BX377759 BX377759
	39	271.6	10.9	2097	4	CR614786	CR614786 full-leng
c	40	271.4	10.9	603	1	AL707313	AL707313 DKFZp686P
c	41	271.2	10.9	5325	4	HSM804049	AL832738 Homo sapi
c	42	271.2	10.9	5785	4	HSM802309	AL137734 Homo sapi
c	43	270.6	10.8	2330	4	CR615928	CR615928 full-leng
	44	270.4	10.8	618	9	BZ609884	BZ609884 WHACH76TF
	45	270	10.8	665	5	BU633001	BU633001 UI-H-DF0-

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS BU785040 507 bp mRNA linear EST 11-OCT-2002
DEFINITION in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3', mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES Location/Qualifiers
source 1. .507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 20.2%; Score 505.4; DB 5; Length 507;
Best Local Similarity 99.8%; Pred. No. 2.4e-35;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1506 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 1565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 448

Qy       1566 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 1625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 388

Qy       1626 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 1685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 328

Qy       1686 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 1745
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 268

Qy       1746 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 208

Qy       1806 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 148

Qy       1866 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCT 1925
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        147 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCT 88

Qy       1926 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 28

Qy       1986 CGACACAGCGAGACTCTATCTCAAAAA 2012
             |||||||||||||||||||||||||||
Db          27 CGACACAGCGAGACTCTATCTCAAAAA 1
```

RESULT 2
BC024593
LOCUS BC024593 4087 bp mRNA linear HTC 29-JUN-2004
DEFINITION Homo sapiens cDNA clone IMAGE:3914314, with apparent retained intron.
ACCESSION BC024593
VERSION BC024593.1 GI:22137609
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4087)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 4087)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 22 Row: m Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.

FEATURES Location/Qualifiers
source 1. .4087
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3914314"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH_MGC_71"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

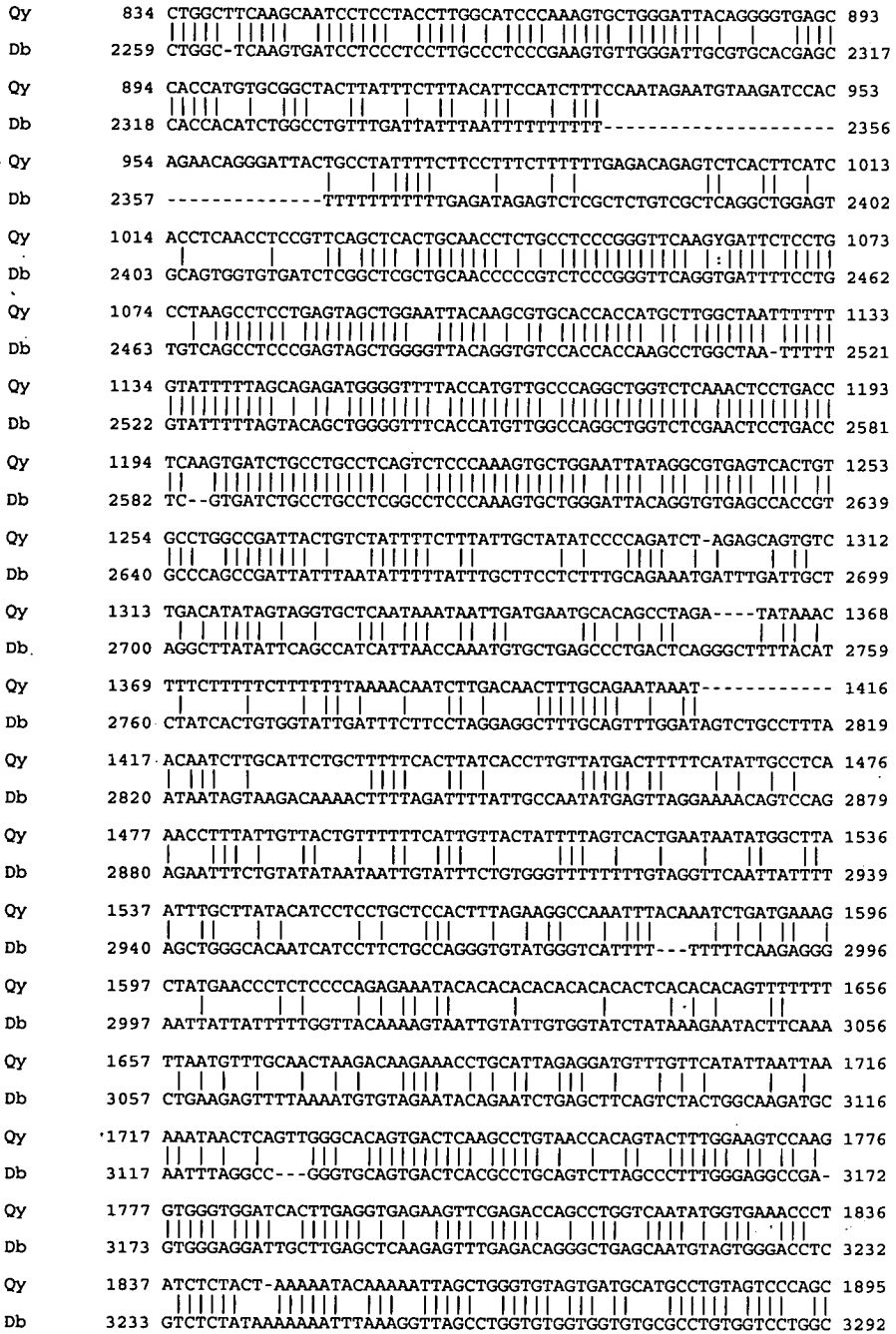
Query Match 13.8%; Score 344.2; DB 4; Length 4087;
Best Local Similarity 58.3%; Pred. No. 6.8e-22;
Matches 859; Conservative 1; Mismatches 544; Indels 69; Gaps 12;

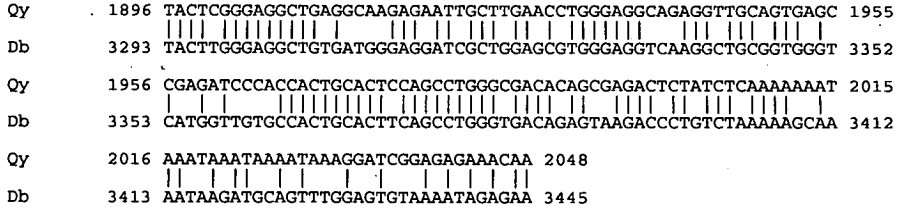
```
Qy        599 TTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 658
              || |   ||| ||||||  ||||| ||||||||||| | |||| |||||||||  |||
Db       2019 TTTGAAATTTGTTTGTTTGTGTTTTGAGACAGGGTCTGGGTCTGCCACCCAGGCTGGAGT 2078

Qy        659 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 718
              |||||||  | ||||| ||| |||||||| ||||||| | | || |||| |||   || |
Db       2079 GTAGTGGTGTGATCATAGCTCAATGCAGCCCCAACCTTCCGGGCGCAAGTGATGCTCTCA 2138

Qy        719 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 778
              ||||||||||  ||||||||||||||  || ||  |||||||| ||   |||||  | |
Db       2139 CCTCAGCCTCCTGAGTAGCTGGGACTACAGGCATGTGCCACCACACCCTCCTAATATATA 2198

Qy        779 AAAT-----TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 833
                  |     ||||||||||||||   ||| | | || |||| |||| || ||||  |||||
Db       2199 TTTTTTTTATTTTTTGTAGAGATGAGGTTTCATTATGTTGCCCAGACTGGTCTGGAACTC 2258
```





```
RESULT 3
HSM803026
LOCUS       HSM803026              3990 bp    mRNA    linear   HTC 16-APR-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp761O0217 (from clone DKFZp761O0217).
ACCESSION   AL713681
VERSION     AL713681.1  GI:19584382
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3990)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp761O0217) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp761O0217
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3990
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp761O0217"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761O0217"
                     /tissue_type="amygdala"
                     /clone_lib="761 (synonym: hamy2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="cytochrome c"
     gene            1. .3990
                     /gene="DKFZp761O0217"
     CDS             61. .378
                     /gene="DKFZp761O0217"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD28485.1"
                     /db_xref="GI:19584383"
                     /db_xref="GOA:P99999"
                     /db_xref="InterPro:IPR002327"
                     /db_xref="InterPro:IPR003088"
                     /db_xref="UniProt/Swiss-Prot:P99999"
                     /translation="MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQA
                     PGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKA
                     TNE"
ORIGIN
```

```
  Query Match           13.6%;  Score 341;  DB 4;  Length 3990;
  Best Local Similarity   57.5%;  Pred. No. 1.3e-21;
  Matches  856;  Conservative    1;  Mismatches  541;  Indels   91;  Gaps   10;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
Db       2522 TTTTCTCATTTATTTTTTCTGAGACGGAGTTTCGCCCTTTTCGCCCAGGCTGAGTACAGT 2581

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
Db       2582 GGTGCAATC-----TCACTGCAACCTCCGCCTTCCAGTTGCAAGTGATTCTCCTGCCTCA 2636

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db       2637 GCCTCCTGAGTAGCTGGGATTGCAGGCACCCGCCACCACGCCTGGCTAATTTTTGTA--T 2694

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db       2695 TTTTAGTGGAGATGGTGTTTCACCATGTTGGCCAGACTGGTCTTGGACTCCTGACCTC-- 2752

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db       2753 GTGATCCACCCACCTTGGCCTCCCAAAGTTCTAGGATTACAGGTGTGAGCCACCACGTCC 2812

Qy        904 GGCTACTTATT---------------------------------------------------- 914
Db       2813 GGCCAATTTTTCTCATTTCTATGCCTCCTATATTAAGGTCTGTGTTGGCACAGATGAGTA 2872

Qy        915 ------TCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACT 968
Db       2873 ACTGCCATGTTCTAGGTCAGTTATACCCAAGCACTTCTGGTGGTTTAAAATGTGATTCTG 2932

Qy        969 GCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC------------ 1016
Db       2933 TAACTTTTTTATTTTTATTTTTTTGAGATAATTTCACTCTTGTTGCCCAGGCTGGAGTGC 2992

Qy       1017 -TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCC 1075
Db       2993 AATGGCGTGATCGCTGCTCACCGCAACCTCCGCCTTCCAGGTTCAAGCGATTCTCCTGAC 3052

Qy       1076 TAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGT 1135
Db       3053 TCAGCCTCTCAAGTAGCTGGGATTACAAGCATGCGCCACCATGCCCAGCTTATTT--TGT 3110

Qy       1136 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195
Db       3111 GTTTTTAGTAGAGACAGGGTTTCTCCATGCTGGACAGGCTGGTCTTGAACTCCCGACGTC 3170

Qy       1196 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1255
Db       3171 GGGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGC 3230

Qy       1256 CTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGA 1315
Db       3231 CTGGCCAATTA---TGTAATTTTTTAAAAAGGACATTTCTATCAGGGATATATACCTTCA 3287

Qy       1316 CATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTT 1375
Db       3288 GAAATAAGGAAATAGGGGAAAAAAAGAGCACTATAAACCACATGTTTTCATTTCTAGTGC 3347

Qy       1376 TTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTT 1435
Db       3348 TTCGCTGTAAGTGGCTAGGTTGGTAGAATCAAAAACAAGGGCCAGATGTATTTAAGGGGT 3407

Qy       1436 TTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTT 1495
Db       3408 ATTCAGATGCCACCTACATGCTTATTTTGTCTAGAACAGTGCTGTCTAATAGAACTTTCT 3467

Qy       1496 TTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCC 1555
Db       3468 GTGACGATGGATATTTTGTAGACTTT------TGCTGTCCAGTGTGGTAGCCACTAACCA 3521

Qy       1556 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1615
```

```
Db       3522 CATGTGGCTGTTAAGCCCTTGAAATATAGCTAGTGTGACTAGAAAGTATTTTATTTTAAA 3581

Qy       1616 AGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGA 1675
Db       3582 TTTACATAGGCACAAGTGGCTAGTGGCTACTGTATTGACATTCTGGGTCTAGGACTAGAA 3641

Qy       1676 CAAGAAACCTGCATTAGAGG--ATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGC 1733
Db       3642 CCAGTGGTCTGTAACAAAAGTACTTTCTCTTTACTCTATTAAATCTAGAATTAGCCGGGC 3701

Qy       1734 ACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGA 1793
Db       3702 ATGGTCGCTCATGCCTGTGGTCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACTTGA 3761

Qy       1794 GGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC 1853
Db       3762 GGTCGGGCGTTTGAGACCAGCCTGGTCAACATGGCGAAACCCTGTCTCTACAAAAAACAT 3821

Qy       1854 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1913
Db       3822 AAAAGTTAGCCAGGTGTGGTGGTGGGCACCTGTAGTCTCAGCTACTTGGGAGGCTGAGGC 3881

Qy       1914 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
Db       3882 ACAAGAATCACTTGAACTTGGGAGGTGGAGGTTGCAGTGAGCCAGGATTGTGCCACTGCA 3941

Qy       1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAA 2022
Db       3942 CTCAAGCCTGGGTGAC-GAGTGAAACTGTCTCCAAAAAAAAAAAAAAA 3989
```

```
RESULT 4
AQ839852/c
LOCUS       AQ839852               15970 bp    DNA     linear   GSS 31-MAR-2000
DEFINITION  260L13-C56 CITB Homo sapiens genomic clone 260L13, genomic survey
            sequence.
ACCESSION   AQ839852
VERSION     AQ839852.1  GI:6652484
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 15970)
  AUTHORS   Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
            Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
            Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
            Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and
            Landes,G.M.
  TITLE     A 6-Mb high-resolution physical and transcription map encompassing
            the hereditary prostate cancer 1 (HPC1) region
  JOURNAL   Genomics 64 (1), 1-14 (2000)
   PUBMED   10708513
COMMENT     Contact: Carpten JD
            Cancer Genetics Branch
            National Human Genome Research Institute/National Institutes of
            Health
            Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
            Tel: 301 435 5626
            Fax: 301 435 5465
            Email: jdc@nhgri.nih.gov
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .15970
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="260L13"
                     /clone_lib="CITB"
ORIGIN
```

```
  Query Match             13.5%;  Score 337.2;  DB 9;  Length 15970;
  Best Local Similarity   57.8%;  Pred. No. 1.2e-21;
  Matches  905;  Conservative    1;  Mismatches  529;  Indels  131;  Gaps   12;

Qy        598 TTTGGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657
Db       6787 TTTTTATTTTTATTTATTTTTTTTTTGAGGTGGAGTATCGCTATGTCACCCAGGCTGGAG 6728

Qy        658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717
Db       6727 TGCAGTGGCATGATATCAGCTCACTGCAACCTTCACCTGCTAGATTCAAGCAGTTCTCGT 6668

Qy        718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 777
Db       6667 GACTCAGCCTCCTGAGTAGCTGTGATTACAGGCACATGCCACCATGCTCAGCTAACT-TT 6609

Qy        778 AAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 837
Db       6608 TGTATTTTTTAGTAGAGACAGGGTCTCACCATGTTGGCCAGGCTGGTCTGGAACTCCTGA 6549

Qy        838 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 897
Db       6548 CCTCAAGTGATCCGCCCTCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACT 6489

Qy        898 ATGTGCGGCTACTTATTTCTTT--------ACATTCCATCTTTCCAATAGAATGTAAGAT 949
Db       6488 GCGCCCAGCCTCTGAGTTCTTTTAAGTCAGAAACTTGAGCTCTTCAGCTCTGATAAATTT 6429

Qy        950 CCACAGAACAGGGATTACTGCCT--ATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007
Db       6428 GGGGGGCAAGGGGACTAATTTTTTCTTTTTCTTTTTCTTTTTTTTAAGATGGAGTCTTGC 6369

Qy       1008 TTCATCACCTCAACCTCCGTTCA--------------GCTCACTGCAACCTCTGCCTCCC 1053
Db       6368 TGTGTTGCCCAGGGTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACCTCTGCCTCCC 6309

Qy       1054 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 1113
Db       6308 AGGTTCAAGCAATTCTCCTGCCTCAGCTTACTGAGTAGCTGGGACTGCAGGCCTGTGCCA 6249

Qy       1114 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 1173
Db       6248 CCACTCTCAGCTAA-TTTTTGTATTTTTAGTAGAGACAGGGTTTCAGCACATTGGCCAGG 6190

Qy       1174 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1233
Db       6189 CTGGTCTCGATCTCCTGACCTCAAGTGATCTGCCTACCTCGTCTCCCCAAAGTGCTGGGA 6130

Qy       1234 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1293
Db       6129 TTACAGGCAAGAGCCACCACACCTGGCCCTT---GGGGGATGTTATTTCTTTGACAAGTT 6073

Qy       1294 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1353
Db       6072 TTGCCTTTCAAATTATATCTGTTGTCTCTTTCAGGAACTCTGTTTAGTTATATTTTGGGT 6013

Qy       1354 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1413
Db       6012 CTTCTAGATTAATCCTTTAATTTTTAAAATATCTATACGGTTCATCTCTTTGGCAATTA 5953

Qy       1414 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1473
Db       5952 GTTCTACTTTATACTTTTTCCTTAATTTTTATTTTCCAACTCTTATTTAAATTTCTGTCA 5893

Qy       1474 TCAAACCTTTATT-----------GTTACTGTTTTTTCATTGTTACTATTTTAGTC---- 1518
Db       5892 TATTTTGTTCATTTCTAAGAGTTATTTCATATTTTTTCACTGTTCCTTTTTTTTTCTTTA 5833

Qy       1519 ---------ACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGA 1569
Db       5832 GGCTAGTCAAGTGATTATTGTTCCTTTTTTAATAGTGTCATATTGTTTCAGGGATACAAA 5773

Qy       1570 A------------------------------GGCCAAATTTACAAATCTGATGAAAGC 1597
```

```
Db       5772 ATCTCTTACCTTTCTAAGAATTGATTATACTTGGGTATATTTTATATTTAGGGGAGAATA 5713

Qy       1598 TATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTT 1657
Db       5712 GGGGTTTTTTGTTTGTTTTGTATTGTTGCTGTTGTTTTGTTGTTCAGTTCTCCTGTTTTT 5653

Qy       1658 TAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAA 1717
Db       5652 TAACTTTTTCTGTTTTGTAATTTTGTTCTCTGTCTGTCATGGTAGTTCTCAAATGTTTTG 5593

Qy       1718 AATAACTCAGTTGGGCACAGTGACTCA---------------------------------- 1744
Db       5592 TGGTCTTCGGTTGTCCACAGTGAAAAATTGTTTTAAAGCACTTGGGGCAGAGCTTATAAA 5533

Qy       1745 ----------AGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAG 1794
Db       5532 CCAACGGATTTGGCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCA--TGAG 5475

Qy       1795 GTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC- 1853
Db       5474 GTCAGGAGATCGAGACCATCCTGGCTAACACAGTGAAACCCTGTCTCTACTAAAAATACA 5415

Qy       1854 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1913
Db       5414 AAAAATTAGCCGGGCGTGGTGGCGGGCACCAATAGTCCCAGCTACTCGGGAGGCTGAGGC 5355

Qy       1914 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
Db       5354 AGGAGAATGGCGTGAACCTGGGAGGCGGAGCTTGCAGTGAGCCAAGATCGCACCACTGCA 5295

Qy       1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGG 2033
Db       5294 CTTGAGCCTAGGCAACAAAGCGAGACTGTCTCAAAAAAAAAATAAAAATAAAAATAAACC 5235

Qy       2034 ATCGGA 2039
Db       5234 AATGGA 5229
```

```
RESULT 5
CD557847/c
LOCUS       CD557847                 897 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  AGENCOURT_14423258 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30390116 5', mRNA sequence.
ACCESSION   CD557847
VERSION     CD557847.1  GI:31583915
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM464  row: k  column: 21
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .897
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30390116"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match            12.6%;  Score 315.8;  DB 6;  Length 897;
  Best Local Similarity  68.1%;  Pred. No. 5.2e-19;
  Matches  469;  Conservative    1;  Mismatches  213;  Indels    6;  Gaps    2;

Qy        583 TGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTG 642
              |||| |       |  ||   || |||| | ||  | ||| |||||| |||||  |||||
Db        765 TGTATGCATCCCGATATTATTTTATTTTATTTTTATTTTTTGAGACATGGTCTCACTCTG 706

Qy        643 TCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGC 702
              |  ||||| |  |||  ||||||    | |  || |||||||| |||| ||||||   |
Db        705 TTGCCCAGCCTGGAGTGCAGTGGCAGGATCTCAGCTCACTGCAACCTCCACCTCCCAGGT 646

Qy        703 TCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCAC 762
              |||||  ||   |  |||||||||||||||||||||||||| |||  | ||||||||||||
Db        645 TCAAGCAATTCTCCCACCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCAC 586

Qy        763 GCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTT 822
              |||  |||||||         ||||  ||| |||  |||| || | | | ||||||||||
Db        585 GCCCAGCTAATTTTTTTGTATTTTAAGTACAGATGGGGTTTTGCCATGCTGGCCAGGCTC 526

Qy        823 GTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTA 882
              ||| ||| ||||| | |||||  |||  || |||||||| ||||||||||||||||||
Db        525 CTCTCAAATTCCTGACCTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGCTGGGATTA 466

Qy        883 CAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAAT 942
              |||| ||||||||||| ||  | ||| | |  |  | |  | | | |      | | || |
Db        465 CAGGCGTGAGCCACCGTGCCCAGCTGCATCCTGATATTTAATGCTTGCCAGAATTTGATT 406

Qy        943 GTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGT 1002
              | |  || |  |           |  |   | |||| ||     |       | |  |     |
Db        405 GCACCATGCTTATTGGCTTCTATCTTTTTTTTTTTTTTTGAGATGGAGACTCACTCTCTCT 346

Qy       1003 CTCACTTCATCACCTCAACCTCCGT----TCAGCTCACTGCAACCTCTGCCTCCCGGGTT 1058
              ||| |   |   | |     ||    |||||||||||||||||  | ||| |||||
Db        345 CTCCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCCACTTCCTGGGTT 286

Qy       1059 CAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATG 1118
              ||||:|||||||||| || ||||||| |||||||||| | ||||   |||  ||||||
Db        285 CAAGCGATTCTCCTGTCTCAGCCTCCAGAGTAGCTGGGACTACAGATGTGTGCCACCACA 226

Qy       1119 CTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGT 1178
              | |||||||  ||| |||||||||| |||||||||||| |||||||| |||| ||   |
Db        225 CCTGGCTAA--TTTCGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGACTAAT 168

Qy       1179 CTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATA 1238
              ||| |||||||||||||| ||||||  ||||||||   | ||||||||||| |||  ||| |
Db        167 CTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGTTGGGGTTACA 108

Qy       1239 GGCGTGAGTCACTGTGCCTGGCCGATTAC 1267
              ||||||||  |||    ||  | | | || |
Db        107 GGCGTGAGCCACCACCCCAGCCGGCTTCC 79


RESULT 6
CR858326
LOCUS       CR858326               3095 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp469P012 (from clone DKFZp469P012).
ACCESSION   CR858326
VERSION     CR858326.1  GI:55727613
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 3095)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
            Osanger,A., Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp469P012) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp469P012
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3095
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp469P012"
                     /tissue_type="kidney"
                     /clone_lib="469 (synonym: pkid1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="hypothetical protein (Homo sapiens)"
     gene            1. .3095
                     /gene="DKFZp469P012"
     CDS             120. .533
                     /gene="DKFZp469P012"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH90562.1"
                     /db_xref="GI:55727614"
                     /translation="MLSGRLVPGLVSMAGRVCLCQGSAGSGAVGPVEAAIRTKLEQAL
                     SPEVLELRNESGGHAVPPGSETHFRVAVVSSRFEGLSPLQRHRLIHAALAEELAGPVH
                     ALAIQARTPAQWGENSQLDTSPPCLGGNKKTLGTP"
ORIGIN

  Query Match            12.4%;  Score 311;  DB 4;  Length 3095;
  Best Local Similarity  56.6%;  Pred. No. 6.2e-19;
  Matches  883;  Conservative    1;  Mismatches  511;  Indels  166;  Gaps   10;

Qy        584 GTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGT 643
              || ||     ||   | | |||||| |||  | |  |||  ||||  | ||||  || |||
Db       1570 GTGAGCCACTACCTGTAGCGTTTTTGTTTTGTTTGTTTTTGAGATGGAGTCTCACTTTGT 1629

Qy        644 CACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCT 703
                ||||  |  |||  ||||||||| | |  || |||||||||||||  |||||   | |
Db       1630 TGCCCAACCTGGAGTGCAGTGGTGCGATCTCAGCTCACTGCAGCCTCCGCCTCCCAGGTT 1689

Qy        704 CAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACG 763
              ||||  ||   |  |||||||||||||||||||||||||||||||   || ||  |||||||
Db       1690 CAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTAGAGGCTTGTGCCACCACA 1749

Qy        764 CCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTG 823
              ||  |||||||         | |||| | ||||| |||| | || | |||| ||||||| |
Db       1750 CCCAGCTAATT---TTTTGTATTTTATGGAGACAGGGTTTCACCATGTTGCCCAGGCTGG 1806

Qy        824 TCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTAC 883
              ||||| |||||||   ||| ||| ||| || ||||  || ||||||  ||||| |||||||
Db       1807 TCTTAGACTCCTGAGCTCAGGCACTCCGCCCACCTCAGCCTCCCAAGGTGCTAGGATTAC 1866

Qy        884 AGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC---------------- 925
```

```
Db       1867 AGGCGTAAGCCACCATGCCCTGCAATGTAGCCCTTTTTGAGCATTACAGATATTGCCATA 1926
Qy        926 ----------------------------------CATCTTTCCAAT------------A 938
Db       1927 TTTCTGAGGGGAGCTCAAGTAATGTCTATAGAAGCAACTTTCCAATAATCTCAATGTGGA 1986
Qy        939 GAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACA 998
Db       1987 AATTATTGGAGAACAAAACTTGGTTAACATACCTTTTTTTTTTTTTTTTTTTTGAGACG 2046
Qy        999 GAGT---------------CTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACC 1043
Db       2047 GAGTTTTGCTCTTGTTGCCCAGACTGGAGTGCAATGGCCCAATCTCAGCTCACTGCAACC 2106
Qy       1044 TCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAA 1103
Db       2107 TCCGCCTCCCAGGTTCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACAG 2166
Qy       1104 GCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCAT 1163
Db       2167 GCTCCTGCCACCATGCCTGGCTAA-TTTTTGTATTTTTAGTGGAGACAAGGTTTCATCAT 2225
Qy       1164 GTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAA 1223
Db       2226 ATTGGTCAGGCTGGTCTCAAACTCCTGACCTCAAGCGATCCGCCTGCGTCAACCTTCCAA 2285
Qy       1224 AGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC-GATTACTGTCTATTTTCTTTA 1282
Db       2286 GTTGCTGGGATTACATGCGTGAGCCACCACACCCGGCCTGCTTAACATATATCTTGACCA 2345
Qy       1283 TTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATT 1342
Db       2346 TTTCTAGCTGAAGAATTTAAGGTGCATGTGGTTGGGCAGTAGGGAGTCATTGACAAATTT 2405
Qy       1343 GATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAAC 1402
Db       2406 TGAGTACAGACATGCCAGAGAGCA-----------TTTGGGGGCAGATGGATCTGGCAGT 2454
Qy       1403 TTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTT 1462
Db       2455 GATGTGTGAGTGAATTGGCATGGGAGGTTGGAGGAACAGAGATCAATCCAGAGTGAAGGT 2514
Qy       1463 TTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTG 1522
Db       2515 AGAACCCTGACTTACAAGTCATCTGTGGGAATGGAAAAGAAGGAGTTTATTTGAGAGCTA 2574
Qy       1523 AATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTAC 1582
Db       2575 CCATAAAGTCAGAAAATGGTTTGATTGGAACAGAAATGGACTAGAGAGGGGAGTCAATGG 2634
Qy       1583 AAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCA 1642
Db       2635 TA-------------------------------------------------------ACCCC 2641
Qy       1643 CACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTG 1702
Db       2642 AAGGCAGTGGTTTTCAGATTTTCTCAGGTAAGAGCTCCCATAGGAGATTG---------- 2691
Qy       1703 TTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTAC 1762
Db       2692 ------TTTAAAATGAAGCTTCTGGCTGGGTGTGGTGGCTCACACCTGTAATCCCAGCAG 2745
Qy       1763 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAA 1822
Db       2746 TTTGGGAGGCCATGGCAGGCGGATTGCGTGAGGTCAGGAGTTCGAGACCAGCTTGGGCAG 2805
Qy       1823 TATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGC 1882
Db       2806 CATGGTGAAACCGTATCTC-ACTAAAAATACAAAAATTAGTTGGGTGTGGTGGCAGCCAC 2864
Qy       1883 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 1942
Db       2865 CTGTAATCCCAGCCTCTTAGGCAGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGA 2924
Qy       1943 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCT 2002
Db       2925 GGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGCAACAAAGTGAAACTCT 2984
Qy       2003 ATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTC 2062
Db       2985 GTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAACCACGTAGATTCAAGGTCCCTATCCTTC 3044
Qy       2063 C 2063
Db       3045 C 3045
```

```
RESULT 7
CR749233/c
LOCUS       CR749233               5797 bp    mRNA    linear   HTC 19-AUG-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686C0331 (from clone DKFZp686C0331).
ACCESSION   CR749233
VERSION     CR749233.1  GI:51476197
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 5797)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp686C0331) is available at
            the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
            Berlin, Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686C0331
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .5797
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686C0331"
                     /tissue_type="testis"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
                     /note="hypothetical protein"
     gene            1. .5797
                     /gene="DKFZp686C0331"
     CDS             103. .1725
                     /gene="DKFZp686C0331"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH18089.1"
                     /db_xref="GI:51476198"
                     /translation="MGPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLG
                     ITVSKPDLITCLEQGRKPLTMKRNEMIAKPSVMCSHFAQDLWPEQSMKDSFQKVVLRR
                     YEKCEHDNLQLKKGCISVDECKVHKEGYNELNQCLTTTPRKICQCDKYVKVLHQFPNS
                     NGQKRGHTGKKPFKYIECGKAFKQFSTLTTHKKIHTGGKPYKCEECGKAFNHSCSLTR
                     HKKIHTGEKPYKCEECGKAFKHSSTLTTHKRNHTGEKPYKCDKCGKAFMSSSTLSKHE
                     IIHTEKKPYKCEECGKAFNRSSTLTTHKIIHTGEKPYKCEECDKAFKYSYTLTTHKRI
                     HTEDKPYKCEECGKAFKYSSTLTTHKRIHTGEKPYKCEECGKAFKRSSDLTTHKIIHT
                     GEKPYKCEECGKAFKYSSNLTTHKKIHTGERPYKCEECGKAFNQSSILTTHKRIHTGE
                     KFYKCEECGKAFKRSSHLTTHKIIHTGEKFYKCEECDKAFKHFSNLTTHKKIHTGKKP
```

```
                    YRCEECGKAFKHSSKLTTHKRFHTGEKPYRCE"
ORIGIN

  Query Match           12.3%;  Score 307.2;  DB 4;  Length 5797;
  Best Local Similarity  56.9%;  Pred. No. 9e-19;
  Matches  902;  Conservative    1;  Mismatches  569;  Indels  114;  Gaps   14;

Qy        605 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 664
              |||||| | || || ||| | ||||| ||||||||||||  |||||||  |||  |||||
Db       4976 TTTTTTCTTTTTTTTTTTTGTGACAGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTG 4917

Qy        665 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 724
              |  | | |    | ||||||||  ||||  |||||||  | |   |     |||  | | ||||||
Db       4916 GCACGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTACACCATTCTCCCG-CCTCAG 4858

Qy        725 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 784
              ||||   ||||||||||||||||   |||  ||||||||||||  |||||||  |       ||
Db       4857 CCTCTGGAGTAGCTGGGACTACAGGCGCCCACCACCACGCCCAGCTAATTTA------TT 4804

Qy        785 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 844
              ||| ||||||||  |||| | |    ||| ||||||  | |||||  | ||||||  | ||| |
Db       4803 TTTAGTAGAGACAGGGTTTCAACGTGTTAGCCAGGATGGTCTTGATCTCCTGACCTCATG 4744

Qy        845 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 904
                ||||  ||| ||| ||| |||||||||||||||||||||||||  |||||||||   |     |
Db       4743 --ATCCACCTGCCTGGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTCCCTG 4686

Qy        905 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 964
              |||  ||  ||  ||                                | | ||||  |  | |     |
Db       4685 GCTTTTTTCTTTCTT----------------------TTTGAGATGGAGTGTCCCTCTGT 4648

Qy        965 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 1024
              ||      |    | |  |   |  | |
Db       4647 CACCCAGGCTGGAGTGCAGTGGTGCTAT------------------------------- 4620

Qy       1025 CGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 1084
                |||||  |||||  ||||||  ||||||  ||||||||: ||||||||||||  || |||
Db       4619 --CTCAGCACACTGAAACCTCCACCTCCCAGGTTCAAGCCATTCTCCTGCCTCGGCTTCC 4562

Qy       1085 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 1144
              ||||||||  | |  |||  | ||||||||||||||     |||||||||  | ||||
Db       4561 CGAGTAGCTAGGAATAC--GGGTGCACCACCATGCCCAACTAATTTTTGTTGTTTT---- 4508

Qy       1145 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1204
                | | |||| || ||||||  ||||||||||  ||||||||||||||  ||||||
Db       4507 -AGTAGAGAGTTTCACAATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCA--TGATCT 4451

Qy       1205 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGAT 1264
              ||| | |      || |||||||||||| |||| |||| |||| ||||| | |  ||| |
Db       4450 GCCCGACGTCACCTACCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGTCCAGCCCA- 4392

Qy       1265 TACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTA 1324
                | |  | | | |||| || |||          ||   |  |       ||    |||
Db       4391 -----TATTCTATTTATATTTCTGTATAATTTTTATTATGACCATAAAAATCACCCTGTA 4337

Qy       1325 GGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTT 1384
              |    | |  | ||||||   | |  | ||  ||  || ||    |    || |  | ||
Db       4336 GTCAATTACAATTTAATTGGACATTTAAAAG--TAACTAAAAGTGTATAATTACACTGTT 4279

Qy       1385 TAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTA 1444
              |  || |      || || |         | |      | |||   | |     | | | |
Db       4278 TGGAATAGAAAGGATAAATGCTAGAGGTGATAGATACCTTATTTACCCTAATGTAATTAC 4219

Qy       1445 TCACCTTGTTATGACTTTTTCATATTG-----CCTCAAACCTTTATTGTTACTGTTTTTT 1499
              |      | |||| ||   ||| | |         | | ||| | |||     |   | | |
Db       4218 TACATATTGTATGCCTGAATCAAAATATTCTCTATAAGACATATATACACATACTATATA 4159

Qy       1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTT---AATTTGCTTATACATCCTCCT 1556
              |    |   | || | | | || || |  | ||   |||||   ||      |
Db       4158 CTCACAAATAGTAATAATAAATTAAGAAAAAAGATTAAAAATTTAACTAACGGGAACAAT 4099

Qy       1557 GCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGA 1616
                ||  | || | ||   |     | |    | |  |  | |   |  || || |  |
```

```
Db       4098 ATTCTTCATTTGCAGTTTACTGGCAAAGTGATTACTAGTGATGACATTCCACTACATACC 4039

Qy       1617 GAAATACACACACACACACACTCACACACAGTTTTTTTTTAATG-------------- 1662
                ||    | ||      |   |  |  |  |||   |||   |||  |
Db       4038 AAATAGTATACTTCTTCCATCTTTTGCTTACACCATTTGAGTAAGGCCGGATAGGTTAAA 3979

Qy       1663 ------TTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAA 1716
                    | ||  | |       |  |||| | ||  |      |    |  ||| ||
Db       3978 GTTAGTGGCGTAATAATGCTTCATTAAATGCACAATAGTTTTAAGATGGCAAAAAATAAA 3919

Qy       1717 AAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAG 1776
              | | || |  | ||| || | || ||||  ||| ||| | ||  ||||||  || || ||
Db       3918 ATAAAATTAGGCTGGACATATTGGCTCACACCTATAATCCCAACACTTTGAGAGCCCGAG 3859

Qy       1777 GTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCT 1836
              |||||||||||  ||||||  | ||  ||||  |||||  |||||||  ||||||||  |||||||
Db       3858 GTGGGTGGATCGCTTGAGCTCAGGAGTTTGAGACTAGCCTGACCAATATGGGGAAACCCC 3799

Qy       1837 ATCTCTACTAAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCT 1896
              || |||| |||||||| ||||||||||   || || |||    |   ||||||||||  ||
Db       3798 ATGTCTATTAAAAATATAAAAATTAGCCAGGCGTGGTGTCACGCACCTGTAGTCCCCACT 3739

Qy       1897 ACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAG----TG 1952
              |||||||||||||||||  |||||||  |||||||  |||||||  ||||||||||     ||
Db       3738 ACTCGGGAGGCTGAGGCAGGAGAATTTCTTGAACTTGGGAGGTGGAGGTTGCAGTCATTG 3679

Qy       1953 AGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAA 2012
                || ||||   |||  ||||  ||||||||||||  ||||  ||| ||||||  |||  |||||
Db       3678 TGCTGAGACTGCACTGCTGCCCTCCAGCCTGGGTGACAGAGCAAGACTCCATCCCAAAAG 3619

Qy       2013 AATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAA 2072
              ||  |  |||  ||||  |||       |   ||      | |  |  |       |  |  ||
Db       3618 AAAAGTTAAGTAAATAAAATATAAGTTAAGTTCACACATAATCTAACAATTTTTAAATAT 3559

Qy       2073 GCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTG 2132
                |   | |  ||  ||||  |    | ||||   |  ||    | |   ||||
Db       3558 ATTGCATTTTATTACGTAACAGTAAAATTAGTAAAATAATTGATATACTTTAATTTATAA 3499

Qy       2133 TTTATTTTGGTTATAAAAGTAAACAA 2158
              |||  |||    | |  | | |  | ||
Db       3498 TTTCTTTCTCTCAGTATAATGTAGAA 3473
```

```
RESULT 8
CR860521/c
LOCUS       CR860521               1605 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp468P2419 (from clone DKFZp468P2419).
ACCESSION   CR860521
VERSION     CR860521.1  GI:55731876
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 1605)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Medigenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project.
            This clone (DKFZp468P2419) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468P2419
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
```

```
FEATURES             Location/Qualifiers
     source          1. .1605
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp468P2419"
                     /tissue_type="heart"
                     /clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="unclassified"
ORIGIN

  Query Match             12.1%;  Score 301.2;  DB 4;  Length 1605;
  Best Local Similarity   67.7%;  Pred. No. 6.7e-18;
  Matches  474;  Conservative    1;  Mismatches  204;  Indels   21;  Gaps    3;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657
              |||   |||||||  | |  ||||||  |||||||  ||||||  |||| ||| |||  |||
Db       1602 TTTTTTTTTTTTTGTTTGTTTGTTTTGAGACAGAATCTTGCCTTGTCTCCCGGGCTGGAG 1543

Qy        658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717
              ||||||||||  |  ||  ||||||||  | ||  |  ||||  |   ||  |   ||   |
Db       1542 TGTAGTGGTGCAATCTCAGCTCACTGCAACATCCATCTCCCGGATTCCACTTATTCTCCT 1483

Qy        718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 777
              ||||||||||  ||  ||||||||||  ||  |  |     |||  |  |||   ||  |  |
Db       1482 GCCTCAGCCTCTCAGGTAGCTGGGATTATGGGTGCCTGCCATAATGCCCAACTCAATTTT 1423

Qy        778 AAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 837
              |   ||||  ||||||||  ||||  |  ||  |  ||||||||||||  ||||  ||||||||||
Db       1422 GTA--CTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGG 1365

Qy        838 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 897
              | |||||   ||||  |||  |||||||  ||||||||||||||||||||||  ||||  |
Db       1364 CCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGACCACTG 1305

Qy        898 ATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAA 957
                  | ||   | | ||  | | | |||     | |         | | ||  |
Db       1304 CACCTGGCCTGAATTTCTCCATTCTTCCCA-----CACACCCGCCTCAGGTTCTCCTTCC 1250

Qy        958 CAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCT 1017
                |||   || | |||| ||  ||| ||||||||||||  |||||||   |||||
Db       1249 TGATCATTGACGCTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGCATCTCACTGTCTCACCC 1190

Qy        1018 CAACCTCCGTTCA--------------GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGY 1063
               ||   || ||                |||||||| |||||||  ||||||  |  |||||:
Db       1189 AGACTGGAGTGCAGTAGTGCAATCTTGGCTCACTGAAACCTCTTCCTCCCAGCCTCAAGC 1130

Qy        1064 GATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGG 1123
               ||||||||||  ||  |||||||  ||||||||||  ||||||  |  |||||||||  |
Db       1129 GATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACATGTGTGCACCACTAACACCCA 1070

Qy        1124 CTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAA 1183
               |  ||||||||  ||||||  ||||||||||||||  |||||  |||  ||||  |||||||  |
Db       1069 GATAATTTTTGTACTTTTAGTAGAGATGGGGTTTGACCATATTGGCCAGACTGGTCTTGA 1010

Qy        1184 ACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGT 1243
               |  ||||||||||  ||||||  |||  ||||||  |||  ||||||||||  |||||||||
Db       1009 ATTCCTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTGTTGGGGTTATAGGCGT 950

Qy        1244 GAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTAT 1283
               |||  |||||  ||  ||||   |  |  ||   ||  |  ||
Db        949 GAGCCACTGCACCCGGCCCCCTTCCTTCATCTTAGTCAAT 910


RESULT 9
CR599842
LOCUS       CR599842               1863 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DA002YO06 of Neuroblastoma of Homo
            sapiens (human).
ACCESSION   CR599842
VERSION     CR599842.1  GI:50480649
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1863)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1863)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1863
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA002YO06"
                     /tissue_type="Neuroblastoma"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             12.1%;  Score 301.2;  DB 4;  Length 1863;
  Best Local Similarity   56.9%;  Pred. No. 6.1e-18;
  Matches  811;  Conservative    1;  Mismatches  509;  Indels  105;  Gaps   10;

Qy        624 GAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTG 683
              | | |||  ||||  ||||||  |||||||  |||  |  ||||  |      | | ||||||
Db        496 GGGTCAGAGTCTCACTCTGTTTCCCAGGCTGGAGTTCTGTGGCACCGTGTTGGCTCACTG 555

Qy        684 CAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGAC 743
              || ||||   ||||||  |  ||||  ||   ||  ||||||||||  ||||||||||
Db        556 CAACCTCCGTCTCCTGGGTTCAAGCAATTCTCTTGCCTCAGCCTCCAGAGTAGCTGGGAT 615

Qy        744 TACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCT 803
              |||  | |  |||||||||  ||  ||||||||     |  |||||  |||||  ||  ||||  |
Db        616 TACAGGTGCCCACCACCACACCCGGCTAATTTTTTTATATTTTTAGTAGACACGGGGTTT 675

Qy        804 TACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCA 863
              || | ||||||||||||  |||||  ||||  ||  |  |||  |  ||  |  ||  |||||||
Db        676 CACCATGTTGGCCAGGCCGGTCTTGAACTTGTGACGTCAGGTGATGCACCCGCCTTGGCC 735

Qy        864 TCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACAT 923
              |||||||||||||||||||||||||  |  |||||||  |  |||
Db        736 TCCCAAAGTGCTGGGATTACAGGTCCAGATTTTGAACAAATCTCGCTAGAGATTGAACAG 795

Qy        924 TCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCT 983
              |  |  ||  ||  ||  |  |  |
Db        796 AGCTGCGCCTGGATCCCGCCACAGTGAGGAGACCTGAAGACCAGAGAAAACACAGCAAAC 855

Qy        984 TTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAA-C 1042
              | ||           |  ||  |                 || |||| ||||||
Db        856 GGAGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAGTGTGGCGATCTCGGCTCCCTGCAAGA 915

Qy        1043 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 1102
               |||||||||| | |||| |: ||| |||  ||| |||| |  ||||||||||| ||||
Db         916 ATCTGCCTCCCAGATTCACGCCATTTTCCCACCTCAGCCCCTGGAGTAGCTGGAAGTACA 975

Qy        1103 AGCGTGCACCACCATGCTTGGCTAAT----TTTTTGTATTTTTAGCAGAGATGGGGTTTT 1158
               ||  | ||||||  || ||||||||  ||||||||||||||| |||||  |||||||
Db         976 GGCTCCCGCCACCACGCCTGGCTAATTTTGTTTTTGTATTTTTAGTAGAGACGGGGTTTC 1035
```

```
Qy       1159 ACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCT 1218
              ||  ||||    ||| |||||||| | ||||||||||||  |||||    | ||||  | ||
Db       1036 ACTGTGTTAGGTAGGATGGTCTCGATCTCCTGACCTCA--TGATCCATCCGCCTTGGCCT 1093

Qy       1219 CCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTC 1278
              | ||||||||||| |||| |||| |||| ||| |  ||||||  |             |||
Db       1094 CTCAAAGTGCTGGGATTACAGGCATGAGCCACCGCACCTGGCTG------------TTTT 1141

Qy       1279 TTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAAT 1338
              || ||| ||||           |   | | | |||| |  |  |||  |     |     |
Db       1142 TTCATTTCTATTATGATTTGTTCTGTGGAATGTCAGTGACTTAGCGGTGTTTTTCACATG 1201

Qy       1339 AATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGA 1398
                |  |    || | ||    |      |  |       | |  | | |  ||  ||
Db       1202 TTTATAAAGCTGAATAGTTCTGTATGTCAGCTCAGCACCCCTAGTGTCAGTCATGCTACT 1261

Qy       1399 CAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGA 1458
              |  ||      | ||| ||     | ||     |  |    || | |      |  | | ||
Db       1262 CTCCTGCTGTGCATAGATGGGGGCCTGTCCACACCCAAATCTCCTGGGTTCCCGCTGTCC 1321

Qy       1459 CTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTC 1518
                  |  |   | ||    |  | | |  |         | |||      |  ||  |
Db       1322 GCACCTGAAGGTACCCACCAGATGTGTCTGTCACTGCCCTCCATCCACCTTTCTTGTGGT 1381

Qy       1519 ACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAAT 1578
              || |    | || ||       | || | | |    ||| ||   || ||
Db       1382 TCTCATGTTTCTGCCTC------CCTAGAGACCATTTCTCTCTACAGCAGTAG------- 1428

Qy       1579 TTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACA 1638
                                          | || | |||||||||    || ||||| |
Db       1429 ----------------------------TTTCCAATAGAAATACAATGCCAGACACATA 1459

Qy       1639 CTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATG 1698
              |         ||  |||| | | | |   | | |    |||||  || |  ||||
Db       1460 C---------GTCATTTTAGATTTTCTAGTAGTCACTTTAGAAAAGTGAAAAAGAGGCCA 1510

Qy       1699 TTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACA 1758
                ||                                |||||  ||||  ||||||| | ||
Db       1511 GCTG--------------------------------CAGTGGCTCACACCTGTAATCCCA 1538

Qy       1759 GTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGG 1818
              | |||||||||| |  ||||||||||||||  |||||  ||  ||  |||||||| |||||
Db       1539 GCACTTTGGAAGGCTGAGGTGGGTGGATCAC--GAGGTCAGGAGATCGAGACCATCCTGG 1596

Qy       1819 TCAATATGGTGAAACCCTATCTCTACTAAAAATAC-AAAAATTAGCTGGGTGTAGTGATG 1877
                ||| ||||||||||||  |||||||||||||||| ||||||||||||||  || |||
Db       1597 CCAACATGGTGAAACCCCCTCTCTACTAAAAATACAAAAAATTAGCTGGGCGTGGTGGCA 1656

Qy       1878 CATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAG 1937
                |||||||||||||||||||| |||||||||||  |||||  ||  |||||  || ||
Db       1657 GGCGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAG 1716

Qy       1938 GCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAG 1997
              ||||||  |||||||||||||||||   | |||||||||||||||||||||||| || |||
Db       1717 GCAGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCACTCCAGCCTGGGCGACAGAGGGAG 1776

Qy       1998 ACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGA 2043
              ||||  ||||||||||| ||| ||    |  ||   |  || || |
Db       1777 ACTCCATCTCAAAAAAAAAAAAAAGGTTACAAAGAAACAGGTGAAA 1822


RESULT 10
BQ958903
LOCUS       BQ958903                946 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10035485 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483214
            5', mRNA sequence.
ACCESSION   BQ958903
VERSION     BQ958903.1  GI:22374381
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 946)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2667  row: g  column: 23
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1. .946
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6483214"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_40"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           12.0%;  Score 300.6;  DB 5;  Length 946;
  Best Local Similarity   69.1%;  Pred. No. 1.1e-17;
  Matches  468;  Conservative    1;  Mismatches  197;  Indels   11;  Gaps    4;

Qy        596 AGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATG 655
              || ||   ||||||||| || || ||| ||||||| ||||||||||||||||||||   |
Db         24 AGNTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCTGG 83

Qy        656 AGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGC 715
              ||  ||||||||   | || | |||||||| ||||| |||||||   | ||| |||   |
Db         84 AGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAGCCTCCTGGATTTAAGTGATTCTC 140

Qy        716 TGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTA 775
              |||||||||||||||||||||||||||| |||  | ||||||||||| ||| ||||||||
Db        141 CTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCATGCCCGGCTAATTT 200

Qy        776 AAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCT 835
                  ||||| || ||||| |||| | || | |||||||||||| ||||  |||||||
Db        201 TTTTGTATTTTTAGTGGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCT 260

Qy        836 GGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCA 895
              | | ||  |  |||| ||| |||  || |||||||||||||||||||||| ||||||||
Db        261 GACCTC--GTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCA 318

Qy        896 CCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAG 955
              ||| |   |||  |          ||  |     | |    |          ||      |
Db        319 CCACGCCTGGCCCCAGTCCAAAATATTTAAAGATTGTTTCCTTAGTGTCTTGAAGTTTTG 378

Qy        956 AACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCAC 1015
              |||     | | | ||||     |   | |         |  |  ||  |   |
Db        379 CACA-----AAATTCTTTTTTTTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGTGC 433

Qy       1016 CTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCC 1075
                | |    |  ||||||||||||||||||||||| |||||||||: |||||||  ||
Db        434 AGTGGCGTGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCAATTCTCCCACC 493

Qy       1076 TAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGT 1135
```

```
Db        494 TCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTTGC 552

Qy       1136 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195

Db        553 ATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGGTGGTCTTGAACTCCTGACCTC 612

Qy       1196 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1255

Db        613 AGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGC 672

Qy       1256 CTGGCCGATTACTGTCT 1272

Db        673 TCAGCCGCANAATTCCT 689


RESULT 11
BQ706343/c
LOCUS       BQ706343                 918 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8484983 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301091
            5', mRNA sequence.
ACCESSION   BQ706343
VERSION     BQ706343.1  GI:21845242
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2516  row: g  column: 12
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6301091"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             11.9%;  Score 298;  DB 5;  Length 918;
  Best Local Similarity   68.6%;  Pred. No. 1.8e-17;
  Matches  456;  Conservative    1;  Mismatches  197;  Indels   11;  Gaps    3;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657

Db        666 TTTTGGATTATGATTATTATTCTTTGAAGCCAGGGTCTCGCTCTGTCACNCAGGCCGGAG 607

Qy        658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717

Db        606 TGCAGTGGTGCAATCATGGCTCACTCCAGCCTTTATCTCCTGGGCCCAAGTGATCCTCCA 547

Qy        718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTA------ 771

Db        546 ACCTCAGCCTCCCAAGTAGCTGGGATTATAAGTGCGTGCTACGATGTCTGGCTAATTTTT 487

Qy        772 -ATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAA 830

Db        486 TATTTATAAATATTTTTTGTAGAGACAGGGTCTCGCTTTGTTGCCCAGGCTGGTCTCAAA 427

Qy        831 CTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTG 890

Db        426 TTCCTGGACTCAGGCGATCCTCCTGCTTCAGCCTCCCAAAGTGCTGAGATTACAGGCGTG 367

Qy        891 AGCC---ACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 947

Db        366 AACCTTTCAACTTTAATGCTTTCAGTATTTTCCCATCCCTTGATTGAACATTTGGGTTT 307

Qy        948 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007

Db        306 TAAATTATTTTTGTTGTTGTTGTTTCTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGC 247

Qy       1008 TTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATT 1067

Db        246 TAGAATGCAGGGGTGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGCGTTCAAGCGATT 187

Qy       1068 CTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA 1127

Db        186 CTTGTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCTGCCATGATGCCCAACTAA 127

Qy       1128 TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTC 1187

Db        126 TTTTTTGTATTTTTAGTAGAGACAGTGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTC 67

Qy       1188 CTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCA-AAGTGCTGGAATTATAGGCGTGAG 1246

Db         66 CTGACCTCAAGTGATCCGCGCACCTTGGCCTCCCANAATTGCTGGGATTACAGGTGTGAG 7

Qy       1247 TCACT 1251

Db           6 CCACT 2


RESULT 12
CX780936
LOCUS       CX780936                 736 bp    mRNA    linear   EST 02-MAR-2005
DEFINITION  HESC3_2_B04.b1_A036 NIH_MGC_260 Homo sapiens cDNA clone
            IMAGE:30928759 3', mRNA sequence.
ACCESSION   CX780936
VERSION     CX780936.1  GI:58297726
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: HESC3_2_B04.g1_A036
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: BresaGen, Inc.
             cDNA Library Preparation: Express Genomics, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
            University of Georgia
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
           Plate: NDAM1172  row: c  column: 08
           Seq primer: M13-21 (TGTAAAACGACGGCCAGT)
           High quality sequence stop: 711
           POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30928759"
                     /sex="male"
                     /tissue_type="human embryonic stem cells"
                     /cell_type="human embryonic stem cells"
                     /cell_line="BG01"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="NIH_MGC_260"
                     /note="Vector: pExpress-1; Site_1: NotI; Site_2: EcoRV;
                     RNA obtained from human embryonic stem cells isolated from
                     the inner cell mass of blastocyst stage embryos. Cell line
                     id and NIH Registry designation is BG01. Positive for
                     SSEA3, SSEA4, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and
                     Nanog expression; negative for SSEA1 expression. Passage
                     number 21. cDNA primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. This primary library
                     is non-normalized (normalized primary library is
                     NIH_MGC_261). It was constructed by Express Genomics
                     (Frederick, MD). Sequence ends have been trimmed to
                     exclude vector and regions below Phred quality 16. Note:
                     this is a Mammalian Gene Collection library."
ORIGIN

  Query Match             11.7%;  Score 293.6;  DB 8;  Length 736;
  Best Local Similarity   66.9%;  Pred. No. 5.1e-17;
  Matches  464;  Conservative    1;  Mismatches  220;  Indels    9;  Gaps    3;

Qy        599 TTGGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 658
              ||   ||||||||| || || ||| ||||  | ||||  |||| |  |||||||   ||
Db          1 TTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTCTTGCCCAGGCTGAAGT 60

Qy        659 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 718
              |||||| |  | |   | |||||||| ||||  ||||||| | |||||   |
Db         61 GCAGTGGCGGGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCACTTCTTCTG 120

Qy        719 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 778
              | |||||| ||  |||||||||||| |||  ||   |||||||| |||||| || ||
Db        121 CTTCAGCCCCCTGAGTAGCTGGGATTACAGGCACCCACCACCAGGCCTGGATACTT--TT 178

Qy        779 AAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGC 838
                | |||||    | | |||| || |||  |||||||||| |||| |||||||| |
Db        179 TTGTATTTTTAGTAAAAATGGGGTTTGCTATATTGGCCAGGCTGGTCTCAAACTCCTGAC 238

Qy        839 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA 898
              ||| |  |||| ||| |||  || | ||||||||||||||||||||  ||| ||||||
Db        239 CTCAGGTGATCCACCTGCCTAAGCCTGCCAAAGTGCTGGGATTACAGGCATGAACCACCA 298

Qy        899 TGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAAC 958
              ||  | ||  |||  ||||||  | |           ||     | |  |    |
Db        299 TGCCCAGCCTCTTTATTCTTTTTAATATTATCACAAGTTATTTATTTATTTATTTATTTA 358

Qy        959 AGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTC 1018
                ||| |   |     |     |||    ||  | |  ||     | | ||    | |
Db        359 TTTATTTATTTTTTGAGATGGAGTTTTGCCCTTGTTGCCCAAGCTGGAGTGCAATGGCAC 418

Qy       1019 AACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAA 1078
              | |     |||||||||||||||||  |||||||||||||:||||||||||||| |
Db        419 GATC-----TCAGCTCACTGCAACCTCCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCA 473

Qy       1079 GCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATT 1138
              |||||||||||||||| |||||| ||      ||| ||||| |||||||  |||||| |||
Db        474 GCCTCCTGAGTAGCTGGGATTACAGGCACATGCCATCATGCCTGGCTAA--TTTTGCATT 531

Qy       1139 TTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAG 1198
              ||||| |||||  | ||||  |||||||  |||||||||||  |||||||||||||| |
Db        532 TTTAGTAGAGACAGAGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGG 591

Qy       1199 TGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTG 1258
              |||||||||  |||| | ||||||||||||||| |||||||||  |||| ||| |||||
Db        592 TGATCTGCCCTCCTCGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCGTGCCCA 651

Qy       1259 GCCGATTACTGTCTATTTTCTTTATTGCTATATC 1292
              |||  |  |      | ||  |  || || | ||
Db        652 GCCATTATCACAAGTTATTTATACTTCCTTTCTC 685


RESULT 13
CR613629
LOCUS       CR613629               3092 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI039YN19 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR613629
VERSION     CR613629.1  GI:50494436
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3092)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 3092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .3092
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI039YN19"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             11.6%;  Score 290.6;  DB 4;  Length 3092;
  Best Local Similarity   69.7%;  Pred. No. 3.7e-17;
  Matches  466;  Conservative    1;  Mismatches  190;  Indels   12;  Gaps    5;

Qy        594 ACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCA 653
              | | |||   ||||||||| || || ||| ||||||| ||||||||||||||||||||||
Db       1750 AAATTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCT 1809

Qy        654 TGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCT 713
              |||  ||||||||   | || | |||||||| ||||| |||||||   | ||| |||
Db       1810 GGAGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAGCCTCCTGGATTTAAGTGATTC 1866

Qy        714 GCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAAT 773
              |  |||||||||||||||||||||||| |||  | ||||||||||| ||| |||||||
Db       1867 TCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCATGCCCGGCTAAT 1926

Qy        774 TAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 833
              |      | ||||| || ||||| |||| | || | |||||||||||| ||||  |||||
Db       1927 TTTTTGTA-TTTTTAGTGGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTC 1985
```

```
Qy        834 CTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC 893
              ||| | ||  |  |||| ||| |||  || ||||||||||||||||||||||| ||||||
Db       1986 CTGACCTC--GTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGC 2043

Qy        894 CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC 953
              ||||| |   |||  |          ||  |      | |    |          ||
Db       2044 CACCACGCCTGGCCCCAGTCCAAAATATTTAAAGATTGTTTCCTTAGTGTCTTGAAGTTT 2103

Qy        954 AGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATC 1013
              | |||      | | | ||||     |  | |        |  |  ||  |
Db       2104 TGCACA-----AAATTCTTTTTTTTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGT 2158

Qy       1014 ACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTG 1073
              |     | |    | ||||||||||||||||||||| |||||||||: |||||||
Db       2159 GCAGTGGCGTGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCAATTCTCCCA 2218

Qy       1074 CCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT 1133
              ||| |||||||  |||||||| ||||||  ||||  |||||| | ||| ||| |||||
Db       2219 CCTCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTT 2277

Qy       1134 GTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACC 1193
              | ||||||||  |||| || |||| |||||||| ||||| ||||||  ||||||||||
Db       2278 GCATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGTTGGTCTTGAACTCCTGACC 2337

Qy       1194 TCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGT 1253
              ||| ||||||  |||||||| | |||||||||||||||| |||| |||| ||| || ||
Db       2338 TCAGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGT 2397

Qy       1254 GCCTGGCCG 1262
              ||   ||||
Db       2398 GCTCAGCCG 2406
```

RESULT 14
BQ681302/c
LOCUS BQ681302 855 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8034339 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213772 5', mRNA sequence.
ACCESSION BQ681302
VERSION BQ681302.1 GI:21793981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 855)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2380 row: a column: 05
High quality sequence stop: 649.
FEATURES Location/Qualifiers
source 1. .855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6213772"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 11.5%; Score 288.2; DB 5; Length 855;
Best Local Similarity 69.2%; Pred. No. 1.4e-16;
Matches 471; Conservative 1; Mismatches 161; Indels 48; Gaps 4;

```
Qy        606 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 665
              ||||||| || || |||  |||||||||||| |||||||||||| |  |||  ||||||
Db        855 TTTTTTTTTTTTTTTTTTTAGACAGGGTCTTGATCTGTCACCCAGACTGGAGTGCAGTGG 796

Qy        666 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 725
              ||  | ||||| |||||| | | |||| ||| |  ||||| |  ||       || |||||
Db        795 TGAGATCATAGCTCACTGTAACNTCAAACTCTTCGGCTCAGGCAATACTTCCACTTCAGC 736

Qy        726 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTT 785
              |||||| |||||||||||||| || |    ||||||  ||  ||||| |      || |||
Db        735 TTCCCAAATAGCTGGGACTACAAGTGCATGCCACCATACCCAGCTAAAT-GTTTAAGTTT 677

Qy        786 TTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGC 845
              ||||||||||  ||||||| | | |||   ||||||  ||||||||||||||||||||||
Db        676 TTTGTAGAGATGGGGTCTTGCCATGTTTCNCAGGCTGCTCTTAAACTCCTGGCTTCAAGT 617

Qy        846 AATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGG 905
              ||||||| ||||  || |||||||||||||||| ||||| ||||| ||| |||    ||
Db        616 GATCCTCCCACCTCAGCCTCCCAAAGTGCTGGGATGACAGGTGTGAGACACAATGCCTGG 557

Qy        906 CTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATT 965
              || ||   |  | |  |  | ||||  | | |  ||
Db        556 CTTCTCTCTCTCTCTCTTTTTTTTTCTGAAACAGAGT---------------------- 518

Qy        966 ACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCC 1025
                                  ||| || ||   | |   |   || | ||    ||| |
Db        517 -----------------TTTGTTCTTGTTGCCCAGGCTGCAGTGCAATGGCTCGA----- 480

Qy       1026 GTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCT 1085
              |||||||||||||||||||| ||||||||||||||:||||||||||||| |||||||
Db        479 TTTCAGCTCACTGCAACCTCTACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCC 420

Qy       1086 GAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA----TTTTTTGTATTTTT 1141
              |||||||| |||||| |||    |||||| || ||||| |    ||||||||||||||
Db        419 AAGTAGCTGGGATTACAGGCGCCTGCCACCACGCCTGGCTGACTTTTTTTTTTGTATTTTT 360

Qy       1142 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1201
              || |||||||||||| |||||| || ||||||||  ||  ||||||||| |||| ||||
Db        359 AGTAGAGATGGGGTTTCACCATGGTGGCCAGGCTGACCTTGAACTCCTGATCTCAGGTGA 300

Qy       1202 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1261
              ||  ||  ||| | ||||||||||||||| |||| ||| ||||| |||||||| ||||
Db        299 TCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCCGGCC 240

Qy       1262 GATTACTGTCTATTTTCTTTA 1282
                | ||  |  | ||  | |
Db        239 AGCTTCTTGCCTTCTTACTGA 219
```

RESULT 15
BQ710315/c
LOCUS BQ710315 769 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8351732 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282212 5', mRNA sequence.
ACCESSION BQ710315
VERSION BQ710315.1 GI:21849214
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 769)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2476  row: d  column: 21
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6282212"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match            11.5%;  Score 287.6;  DB 5;  Length 769;
  Best Local Similarity  67.0%;  Pred. No. 1.6e-16;
  Matches  453;  Conservative    1;  Mismatches  215;  Indels    7;  Gaps    3;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
              ||||||||| || || | | ||||||| ||||  ||||||| |||||||  |||  ||||
Db        768 TTTTTTTTTTTTTTTTTATTGAGACAGAGTCTCACTCTGTCCCCCAGGCTGGAGTGCAGT 709

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
              || |  | |   | |||||||| ||||  ||||||| | ||||| |||   |  |||||
Db        708 GGGGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCA 649

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
              |||||| ||||||||||||||   |||  | ||||   ||| |   | ||  | |   |
Db        648 GCCTCCCGAGTAGCTGGGACTATAGGCGCCCGCCACTGTGCCGGCTAATTTTTATA---T 592

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
              |||| ||||||||  |||| |  | | ||||||||||| ||||  || ||||  | |||
Db        591 TTTTAGTAGAGACGGGGTTTCCCCATGTTGGCCAGGCTGGTCTCGAATTCCTAACCTCAG 532

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
              |  |||| |||  ||  || ||||||  |||||||||||||||| ||||||||||||  |
Db        531 GTGATCCGCCTGGCTCAGCCTCCCAACATGCTGGGATTACAGGTGTGAGCCACCATGCCC 472

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
              ||||   |  ||||  |  |||  | |  |  | | ||  |
Db        471 GGCTGTGTTACACTTTTTGTGGTGTCTCACAGGTCTCTGAGGCTCTGCTCATTTTTGTTT 412

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
              ||  |     |||| |         | |           |||  |   |
Db        411 TTGTTTTTATTTTTTTGAGACAGAGTCTCTTTCTGTCACTAGACTGGAGTGCAGTGGTGC 352

Qy       1024 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 1083
                || |||||||||||||||||||||| | |||||||:|||||| ||||| ||||||
Db        351 AATCTCGGCTCACTGCAACCTCTGCCTCCAGAGTTCAAGCGATTCTCTTGCCTCAGCCTC 292

Qy       1084 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA--TTTTTTGTATTTTT 1141
              |  ||||||||| |||||| ||     |||||||||    ||||  ||||||||||||||
Db        291 CCAAGTAGCTGGGATTACAGGCACATGCCACCATGCCCAACTAATTTTTTTTGTATTTTT 232

Qy       1142 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1201
              || |||||||||||||| ||||||||| ||||| ||||||  | ||| |||||||  ||||
Db        231 AGTAGAGATGGGGTTTGACCATGTTGGCCAGGATGGTCTAGATCTCTTGACCTC--GTGA 174

Qy       1202 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1261
              ||    |  |||  | |||||||||||||   |||| ||||||||| ||||| ||||||||
Db        173 TCCAACCACCTTGGCCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACTGCGCCTGGCC 114

Qy       1262 GATTACTGTCTATTTT 1277
              |  | ||   | || |
Db        113 GGCTCCTCCTTCTTAT 98


Search completed: February 10, 2006, 00:58:23
Job time : 9357.29 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 14:43:53 ; Search time 12458.7 Seconds
(without alignments)
11410.963 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 2499.4
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2499.4	100.0	8368	8	AY438977	AY438977 Homo sapi
c	2	2499.4	100.0	122302	8	AC003982	AC003982 Homo sapi
c	3	2499.4	100.0	220384	14	AC078926	AC078926 Homo sapi
	4	2499	100.0	13612	6	AX377239	AX377239 Sequence
	5	2110	84.4	189729	14	AC073930	AC073930 Homo sapi
c	6	462.6	18.5	178887	14	AC068981	AC068981 Homo sapi
	7	462.6	18.5	190162	8	AC009244	AC009244 Homo sapi
	8	460.6	18.4	123192	14	AP000589	AP000589 Homo sapi
	9	460.4	18.4	150266	8	AC022408	AC022408 Homo sapi
	10	459.4	18.4	167996	8	AC021753	AC021753 Homo sapi
	11	457.4	18.3	137693	8	AC003689	AC003689 Homo sapi
	12	456.2	18.3	201460	8	AP003721	AP003721 Homo sapi
	13	451.2	18.1	195986	14	AC116933	AC116933 Papio anu
c	14	449.2	18.0	134210	8	AC005052	AC005052 Homo sapi
c	15	447.8	17.9	131943	8	AC005484	AC005484 Homo sapi
c	16	447.8	17.9	136119	8	HS1028D15	AL121886 Human DNA
	17	447.8	17.9	186197	14	AL627384	AL627384 Homo sapi
	18	446.2	17.9	163681	8	AL136992	AL136992 Human DNA
	19	445	17.8	117336	8	AC093171	AC093171 Homo sapi
	20	442.6	17.7	91927	8	AC004771	AC004771 Homo sapi
c	21	441.8	17.7	200430	8	AC011500	AC011500 Homo sapi
c	22	439.4	17.6	180076	14	AC022177	AC022177 Homo sapi
c	23	438.8	17.6	201227	14	AC160565	AC160565 Pan trogl
c	24	438.6	17.5	163704	14	AC141415	AC141415 Pan trogl
c	25	438.2	17.5	199045	14	AC149081	AC149081 Pan trogl
c	26	437.8	17.5	185574	14	AC161822	AC161822 Pan trogl
c	27	435.8	17.4	203790	8	AC010422	AC010422 Homo sapi
	28	435	17.4	119043	8	AC139451	AC139451 Homo sapi
c	29	435	17.4	163542	14	AC129071	AC129071 Pan trogl
c	30	434.8	17.4	96995	8	AC008149	AC008149 Homo sapi
	31	434.2	17.4	152659	8	AL591503	AL591503 Human DNA
c	32	433.8	17.4	199384	8	AP003531	AP003531 Homo sapi
	33	433.2	17.3	89448	8	AC005067	AC005067 Homo sapi
c	34	433.2	17.3	178976	14	AL391095	AL391095 Homo sapi
	35	432.6	17.3	156933	8	AC020750	AC020750 Homo sapi
c	36	432.2	17.3	81117	8	AC016898	AC016898 Homo sapi
c	37	432.2	17.3	192961	8	AP002770	AP002770 Homo sapi
c	38	432	17.3	153392	8	AL133548	AL133548 Human DNA
	39	432	17.3	235183	14	AC160951	AC160951 Colobus g
c	40	431.4	17.3	214031	14	AC160952	AC160952 Colobus g
c	41	431	17.2	134957	8	HSJ930L11	AL049759 Human DNA
c	42	430.8	17.2	44062	8	AC004517	AC004517 Homo sapi
c	43	430.8	17.2	118396	8	AC073138	AC073138 Homo sapi
	44	430.8	17.2	162701	14	AC073317	AC073317 Homo sapi
c	45	430.6	17.2	41613	8	AC004790	AC004790 Homo sapi

ALIGNMENTS

RESULT 1
AY438977
LOCUS AY438977 8368 bp DNA linear PRI 29-OCT-2003
DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
ACCESSION AY438977
VERSION AY438977.1 GI:37953284
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8368)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).
FEATURES Location/Qualifiers
source 1. .8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101. .257
/rpt_family="MIR"
/rpt_type=dispersed
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326. .538
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 354. .496

```
                /note="Region not scanned for variation"
gene            875. .6548
                /gene="PLA2G1B"
mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                /gene="PLA2G1B"
                /product="phospholipase A2, group IB (pancreas)"
CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                /gene="PLA2G1B"
                /codon_start=1
                /product="phospholipase A2, group IB (pancreas)"
                /protein_id="AAR05441.1"
                /db_xref="GI:37953285"
                /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
repeat_region   1174. .1307
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1286
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="g"
repeat_region   1319. .1588
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1343
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="a"
variation       1593
                /gene="PLA2G1B"
                /frequency="0.02"
                /replace="t"
repeat_region   1605. .1669
                /rpt_family="L2"
                /rpt_type=dispersed
misc_feature    1638. .2188
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
repeat_region   1670. .1975
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   1976. .2038
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2086. .2165
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2194. .2319
                /rpt_family="Alu"
                /rpt_type=dispersed
misc_feature    2319. .2453
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
variation       3027
                /gene="PLA2G1B"
                /frequency="0.25"
                /replace="t"
repeat_region   3165. .3431
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       3262
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3321
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
variation       3396
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="g"
variation       3630
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="c"
variation       3702
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="a"
variation       3768
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3818
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
repeat_region   3846. .3964
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   3977. .4280
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   4281. .4346
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   4347. .4634
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       4386
                /gene="PLA2G1B"
                /frequency="0.08"
                /replace="t"
repeat_region   4635. .4723
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5013
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="a"
repeat_region   5103. .5404
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5133
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5543
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5555
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       5556
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="g"
repeat_region   5577. .5629
                /rpt_family="L1"
                /rpt_type=dispersed
variation       5580
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
variation       5675
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
repeat_region   5779. .6083
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5929
```

```
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             100.0%;  Score 2499.4;  DB 8;  Length 8368;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2497;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3374 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3433

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3434 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3493

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3494 TGAAGCATCTTGGCGAAGTCATCCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3553

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3554 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3613

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       3614 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3673

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3674 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3733

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3734 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 3793

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3794 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 3853

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3854 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 3913

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3914 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 3973

Qy        601 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3974 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 4033

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4034 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 4093

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4094 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 4153

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4154 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 4213

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4214 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 4273

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4274 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 4333

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4334 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 4393

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       4394 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 4453

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4454 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 4513

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4514 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 4573

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4574 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 4633

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4634 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 4693

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4694 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 4753

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4754 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 4813

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4814 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 4873

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4874 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 4933

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4934 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 4993

Qy       1621 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4994 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 5053

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5054 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 5113

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 5173

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 5233

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5234 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 5293

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 5353

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 5413

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 5473

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 5533

Qy       2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 5593

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 5653

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 5713

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 5773

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 5833

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
              |||||||||||||||||||||||||||||||||||||||||
Db       5834 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 5874


RESULT 2
AC003982/c
LOCUS       AC003982               122302 bp    DNA     linear   PRI 18-MAR-1999
DEFINITION  Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION   AC003982
VERSION     AC003982.1  GI:2769695
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122302)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 122302)
  AUTHORS   Bradshaw,H., Wu,X. and Ozersky,P.
  TITLE     The sequence of Homo sapiens PAC clone 166H1
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was originally isolated in the laboratory of Professor
            Graeme Bell, Howard Hughes Medical Institute and Departments of
            Biochemistry and Molecular Biology, and Medicine, The University of
            Chicago, Chicago, IL, USA.  The clone was provided by the
            laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
            Genetics, Oxford, UK.  Some contig information was also obtained
            from Yamagata et al., Nature 384:455-8 (1996).

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is 278C19; the clone sequenced to
            the right is 15E1.  Actual start of this clone is at base position
            1 of 166H1; actual end is at 122302 of 166H1.
FEATURES             Location/Qualifiers
     source          1. .122302
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q"
                     /clone="166H1"
                     /clone_lib="RPCI-1"
     repeat_region   1. .238
                     /rpt_family="Alu"
     repeat_region   239. .265
                     /rpt_family="AT_rich"
     repeat_region   268. .565
                     /rpt_family="Alu"
     repeat_region   1001. .1309
                     /rpt_family="Alu"
     repeat_region   1440. .1745
                     /rpt_family="Alu"
     repeat_region   1909. .1940
                     /rpt_family="MER1_type"
     repeat_region   1955. .1983
                     /rpt_family="AT_rich"
     repeat_region   2063. .2362
                     /rpt_family="Alu"
     repeat_region   2516. .2650
                     /rpt_family="Alu"
     repeat_region   2674. .2730
```

```
                /rpt_family="L1"
repeat_region   2735. .2788
                /rpt_family="(GAAAA)n"
repeat_region   2790. .3077
                /rpt_family="Alu"
repeat_region   3081. .3139
                /rpt_family="L1"
repeat_region   3716. .4020
                /rpt_family="Alu"
repeat_region   4196. .4492
                /rpt_family="Alu"
gene            4560. .14034
                /gene="WUGSC:H_166H1.1"
CDS             join(4560. .5056,13418. .13748,13882. .14034)
                /gene="WUGSC:H_166H1.1"
                /note="unknown function; 60% similar to Z50177
                (PID:g927403) (PID:g927402); H_166H1.1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB95634.1"
                /db_xref="GI:2769696"
                /translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
                KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
                QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
                HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
                LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
                VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
repeat_region   5366. .5655
                /rpt_family="Alu"
repeat_region   5672. .5798
                /rpt_family="Alu"
repeat_region   5802. .5834
                /rpt_family="(TA)n"
repeat_region   5834. .5857
                /rpt_family="POLY_A"
repeat_region   5870. .6184
                /rpt_family="Alu"
repeat_region   6224. .6268
                /rpt_family="(CA)n"
repeat_region   6269. .6558
                /rpt_family="Alu"
repeat_region   6729. .7029
                /rpt_family="Alu"
repeat_region   7051. .7172
                /rpt_family="Alu"
repeat_region   7192. .7500
                /rpt_family="Alu"
repeat_region   7518. .7820
                /rpt_family="Alu"
repeat_region   8207. .8500
                /rpt_family="Alu"
repeat_region   8576. .8878
                /rpt_family="Alu"
repeat_region   9023. .9154
                /rpt_family="Alu"
repeat_region   9165. .9470
                /rpt_family="Alu"
repeat_region   9887. .10188
                /rpt_family="Alu"
repeat_region   10221. .10342
                /rpt_family="L1"
repeat_region   10408. .10707
                /rpt_family="Alu"
repeat_region   10713. .11014
                /rpt_family="Alu"
repeat_region   11020. .11110
                /rpt_family="Alu"
repeat_region   11146. .11386
                /rpt_family="Alu"
repeat_region   11208. .11386
                /rpt_family="Alu"
repeat_region   11391. .11515
                /rpt_family="Alu"
repeat_region   11568. .11705
                /rpt_family="Alu"
repeat_region   11737. .12032
                /rpt_family="Alu"
repeat_region   12034. .12066
                /rpt_family="7SLRNA"
repeat_region   12075. .12372
                /rpt_family="Alu"
repeat_region   12448. .12753
                /rpt_family="Alu"
repeat_region   12763. .12874
                /rpt_family="L2"
repeat_region   12999. .13303
                /rpt_family="Alu"
misc_feature    complement(13674. .13752)
                /note="match to EST N80020 (NID:g1242721) za91a08.s1"
misc_feature    complement(13840. .14195)
                /note="match to EST N80020 (NID:g1242721) za91a08.s1"
repeat_region   14196. .14229
                /rpt_family="AT_rich"
repeat_region   14240. .14339
                /rpt_family="U6"
repeat_region   14349. .14646
                /rpt_family="Alu"
repeat_region   14652. .14940
                /rpt_family="Alu"

  Query Match             100.0%;  Score 2499.4;  DB 8;  Length 122302;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2497;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26288 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATG 26229

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26228 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 26169

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26168 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 26109

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26108 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 26049

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db      26048 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 25989

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25988 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 25929

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25928 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 25869

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25868 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 25809

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25808 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 25749

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25748 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 25689

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      25688 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 25629

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25628 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 25569

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25568 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 25509

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25508 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 25449

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25448 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 25389

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25388 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 25329

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25328 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 25269

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db      25268 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 25209

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25208 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 25149

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25148 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 25089

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25088 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 25029

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25028 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 24969

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24968 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 24909

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24908 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 24849

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24848 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 24789

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
              ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db      24788 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 24729

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24728 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 24669

Qy       1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24668 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 24609

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24608 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 24549

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24548 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 24489

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24488 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 24429

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24428 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 24369

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24368 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 24309

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24308 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 24249

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24248 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 24189

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24188 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 24129

Qy       2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24128 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 24069

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24068 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 24009

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24008 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 23949

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23948 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 23889

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23888 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 23829

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
              |||||||||||||||||||||||||||||||||||||||||
Db      23828 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 23788
```

```
RESULT 3
AC078926/c
LOCUS       AC078926              220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
```

```
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match             100.0%;  Score 2499.4;  DB 14;  Length 220384;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2497;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71511 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 71452

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71451 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 71392

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71391 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 71332
```

```
Qy      181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71331 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 71272

Qy      241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
            ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db    71271 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 71212

Qy      301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
            ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71211 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 71152

Qy      361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71151 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 71092

Qy      421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71091 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 71032

Qy      481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71031 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 70972

Qy      541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70971 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 70912

Qy      601 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70911 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 70852

Qy      661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70851 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 70792

Qy      721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70791 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 70732

Qy      781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70731 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 70672

Qy      841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70671 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 70612

Qy      901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70611 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 70552

Qy      961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70551 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 70492

Qy     1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
            ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db    70491 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 70432

Qy     1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70431 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 70372

Qy     1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70371 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 70312

Qy     1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70311 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 70252

Qy     1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70251 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 70192

Qy     1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70191 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 70132

Qy     1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70131 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 70072

Qy     1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70071 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 70012

Qy     1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70011 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 69952

Qy     1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69951 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 69892

Qy     1621 TACACACACACACACACTCACACACAGTTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69891 TACACACACACACACACTCACACACAGTTTTTTTTTTAATGTTTGCAACTAAGACAAGA 69832

Qy     1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69831 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 69772

Qy     1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69771 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 69712

Qy     1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69711 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 69652

Qy     1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69651 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 69592

Qy     1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69591 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 69532

Qy     1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69531 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 69472

Qy     2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69471 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 69412

Qy     2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69411 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 69352

Qy     2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69351 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 69292

Qy     2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69291 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 69232

Qy     2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69231 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 69172

Qy     2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      69171 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 69112

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69111 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 69052

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
              |||||||||||||||||||||||||||||||||||||||||
Db      69051 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 69011


RESULT 4
AX377239
LOCUS       AX377239               13612 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212562.
ACCESSION   AX377239
VERSION     AX377239.1  GI:19573528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 1 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .13612
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     variation       3845
                     /note="PS1: polymorphic base G or A"
     variation       3968
                     /note="PS2: polymorphic base A or G"
     variation       6060
                     /note="PS2: polymorphic base A or G"
     variation       6844
                     /note="PS4: polymorphic base G or A"
     variation       9531
                     /note="PS5: polymorphic base G or A"
ORIGIN

  Query Match            100.0%;  Score 2499.4;  DB 6;  Length 13612;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2496;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755

Qy        241 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db       6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815

Qy        301 MMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::|||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db       6816 ACCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7116 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7235

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7236 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7295

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7296 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7355

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7356 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7415

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7416 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7475

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7476 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 7535

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db       7536 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 7595

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7596 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 7655

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7656 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 7715

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7716 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 7775

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7776 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 7835

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7836 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 7895

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7896 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 7955

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7956 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8015

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8016 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8075

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8076 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8135

Qy       1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8136 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8195

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8196 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8255

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8256 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8315

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8316 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8375

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8376 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8435

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8436 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8495

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8496 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 8555

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8556 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 8615

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8616 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 8675

Qy       2161 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8676 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 8735

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8736 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 8795

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8796 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 8855

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8856 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 8915

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8916 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 8975

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
                |||||||||||||||||||||||||||||||||||||||||
Db       8976 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9016

RESULT 5
```

```
AC073930
LOCUS       AC073930              189729 bp    DNA     linear   HTG 01-MAR-2002
DEFINITION  Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC073930
VERSION     AC073930.23  GI:18958593
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 189729)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14861669.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
```

```
            ---------- Project Information
                Center project name: HBWS
                Center clone name: RP11-144B2
            ---------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191640 bases at least Q40
                Consensus quality: 195135 bases at least Q30
                Consensus quality: 197786 bases at least Q20
                Estimated insert size: 193340; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64773: contig of 64773 bp in length
            *    64774    64873: gap of unknown length
            *    64874    99097: contig of 34224 bp in length
            *    99098    99197: gap of unknown length
            *    99198   135331: contig of 36134 bp in length
            *   135332   135431: gap of unknown length
            *   135432   161344: contig of 25913 bp in length
            *   161345   161444: gap of unknown length
            *   161445   171231: contig of 9787 bp in length
            *   171232   171331: gap of unknown length
            *   171332   180922: contig of 9591 bp in length
            *   180923   181022: gap of unknown length
            *   181023   187271: contig of 6249 bp in length
            *   187272   187371: gap of unknown length
            *   187372   189729: contig of 2358 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-144B2"
     gap             64774. .64873
                     /estimated_length=unknown
     gap             99098. .99197
                     /estimated_length=unknown
     gap             135332. .135431
                     /estimated_length=unknown
     gap             161345. .161444
                     /estimated_length=unknown
     gap             171232. .171331
                     /estimated_length=unknown
     gap             180923. .181022
                     /estimated_length=unknown
     gap             187272. .187371
                     /estimated_length=unknown
ORIGIN

  Query Match             84.4%;  Score 2110;  DB 14;  Length 189729;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2110;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        390 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171332 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 171391

Qy        450 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     171392 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 171451

Qy        510 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171452 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 171511

Qy        570 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171512 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 171571

Qy        630 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171572 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 171631

Qy        690 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171632 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 171691

Qy        750 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171692 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 171751

Qy        810 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171752 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 171811

Qy        870 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171812 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 171871

Qy        930 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171872 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 171931

Qy        990 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171932 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 171991

Qy       1050 TCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 1109
              ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db     171992 TCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 172051

Qy       1110 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172052 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 172111

Qy       1170 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172112 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 172171

Qy       1230 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172172 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 172231

Qy       1290 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172232 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 172291

Qy       1350 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172292 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 172351

Qy       1410 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172352 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 172411

Qy       1470 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172412 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 172471

Qy       1530 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172472 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 172531
```

```
Qy       1590 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 1649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172532 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 172591

Qy       1650 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1709
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172592 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 172651

Qy       1710 TAATTAAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1769
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172652 TAATTAAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 172711

Qy       1770 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1829
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172712 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 172771

Qy       1830 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172772 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 172831

Qy       1890 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1949
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172832 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 172891

Qy       1950 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 2009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172892 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 172951

Qy       2010 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 2069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172952 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 173011

Qy       2070 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 2129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173012 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 173071

Qy       2130 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACAT 2189
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     173072 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACGCTACAT 173131

Qy       2190 AAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTT 2249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173132 AAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTT 173191

Qy       2250 ACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCAT 2309
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173192 ACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCAT 173251

Qy       2310 GTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCAT 2369
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173252 GTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCAT 173311

Qy       2370 AATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCT 2429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173312 AATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCT 173371

Qy       2430 GTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGAC 2489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173372 GTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGAC 173431

Qy       2490 CAGCCTGGACAA 2501
             ||||||||||||
Db     173432 CAGCCTGGACAA 173443
```

```
RESULT 6
AC068981/c
LOCUS       AC068981              178887 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-580I7, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC068981
VERSION     AC068981.3  GI:9958291
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:8748938.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0580I07
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157045 bases at least Q40
            Consensus quality: 163476 bases at least Q30
            Consensus quality: 167183 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 174828; sum-of-contigs
            Quality coverage: 3.01 in Q20 bases; agarose-fp
            Quality coverage: 3.40 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1242: contig of 1242 bp in length
            *     1243     1342: gap of unknown length
            *     1343     2435: contig of 1093 bp in length
            *     2436     2535: gap of unknown length
            *     2536     3664: contig of 1129 bp in length
            *     3665     3764: gap of unknown length
            *     3765     4864: contig of 1100 bp in length
            *     4865     4964: gap of unknown length
            *     4965     7695: contig of 2731 bp in length
            *     7696     7795: gap of unknown length
            *     7796    10187: contig of 2392 bp in length
            *    10188    10287: gap of unknown length
            *    10288    12303: contig of 2016 bp in length
            *    12304    12403: gap of unknown length
            *    12404    14390: contig of 1987 bp in length
            *    14391    14490: gap of unknown length
            *    14491    16285: contig of 1795 bp in length
            *    16286    16385: gap of unknown length
            *    16386    18364: contig of 1979 bp in length
            *    18365    18464: gap of unknown length
            *    18465    20794: contig of 2330 bp in length
            *    20795    20894: gap of unknown length
            *    20895    23473: contig of 2579 bp in length
            *    23474    23573: gap of unknown length
            *    23574    25387: contig of 1814 bp in length
            *    25388    25487: gap of unknown length
            *    25488    28870: contig of 3383 bp in length
```

```
            *    28871    28970: gap of unknown length
            *    28971    32203: contig of 3233 bp in length
            *    32204    32303: gap of unknown length
            *    32304    32762: contig of 459 bp in length
            *    32763    32862: gap of unknown length
            *    32863    36428: contig of 3566 bp in length
            *    36429    36528: gap of unknown length
            *    36529    39558: contig of 3030 bp in length
            *    39559    39658: gap of unknown length
            *    39659    44249: contig of 4591 bp in length
            *    44250    44349: gap of unknown length
            *    44350    48970: contig of 4621 bp in length
            *    48971    49070: gap of unknown length
            *    49071    54728: contig of 5658 bp in length
            *    54729    54828: gap of unknown length
            *    54829    60549: contig of 5721 bp in length
            *    60550    60649: gap of unknown length
            *    60650    66747: contig of 6098 bp in length
            *    66748    66847: gap of unknown length
            *    66848    73820: contig of 6973 bp in length
            *    73821    73920: gap of unknown length
            *    73921    80722: contig of 6802 bp in length
            *    80723    80822: gap of unknown length
            *    80823    86541: contig of 5719 bp in length
            *    86542    86641: gap of unknown length
            *    86642    91906: contig of 5265 bp in length
            *    91907    92006: gap of unknown length
            *    92007    99976: contig of 7970 bp in length
            *    99977   100076: gap of unknown length
            *   100077   108154: contig of 8078 bp in length
            *   108155   108254: gap of unknown length
            *   108255   114943: contig of 6689 bp in length
            *   114944   115043: gap of unknown length
            *   115044   122680: contig of 7637 bp in length
            *   122681   122780: gap of unknown length
            *   122781   130297: contig of 7517 bp in length
            *   130298   130397: gap of unknown length
            *   130398   138520: contig of 8123 bp in length
            *   138521   138620: gap of unknown length
            *   138621   146448: contig of 7828 bp in length
            *   146449   146548: gap of unknown length
            *   146549   152895: contig of 6347 bp in length
            *   152896   152995: gap of unknown length
            *   152996   162254: contig of 9259 bp in length
            *   162255   162354: gap of unknown length
            *   162355   178887: contig of 16533 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178887
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-580I7"
     misc_feature    1. .1242
                     /note="assembly_name:Contig22"
     gap             1243. .1342
                     /estimated_length=unknown
     misc_feature    1343. .2435
                     /note="assembly_name:Contig24"
     gap             2436. .2535
                     /estimated_length=unknown
     misc_feature    2536. .3664
                     /note="assembly_name:Contig25"
     gap             3665. .3764
                     /estimated_length=unknown
     misc_feature    3765. .4864
                     /note="assembly_name:Contig27"
     gap             4865. .4964
                     /estimated_length=unknown
     misc_feature    4965. .7695
                     /note="assembly_name:Contig28"
     gap             7696. .7795
                     /estimated_length=unknown
     misc_feature    7796. .10187
                     /note="assembly_name:Contig30"
     gap             10188. .10287
                     /estimated_length=unknown
     misc_feature    10288. .12303
                     /note="assembly_name:Contig31"
     gap             12304. .12403
                     /estimated_length=unknown
     misc_feature    12404. .14390
                     /note="assembly_name:Contig32"
     gap             14391. .14490
                     /estimated_length=unknown
     misc_feature    14491. .16285
                     /note="assembly_name:Contig33"
     gap             16286. .16385
                     /estimated_length=unknown
     misc_feature    16386. .18364
                     /note="assembly_name:Contig34"
     gap             18365. .18464
                     /estimated_length=unknown
     misc_feature    18465. .20794
                     /note="assembly_name:Contig35"
     gap             20795. .20894
                     /estimated_length=unknown
     misc_feature    20895. .23473
                     /note="assembly_name:Contig36"
     gap             23474. .23573
                     /estimated_length=unknown
     misc_feature    23574. .25387
                     /note="assembly_name:Contig37"
     gap             25388. .25487
                     /estimated_length=unknown
     misc_feature    25488. .28870
                     /note="assembly_name:Contig38"
     gap             28871. .28970
                     /estimated_length=unknown
     misc_feature    28971. .32203
                     /note="assembly_name:Contig39
                     clone_end:SP6
                     vector_side:right"
     gap             32204. .32303
                     /estimated_length=unknown
     misc_feature    32304. .32762
                     /note="assembly_name:Contig12"
     gap             32763. .32862
                     /estimated_length=unknown
     misc_feature    32863. .36428
                     /note="assembly_name:Contig40"
     gap             36429. .36528

  Query Match             18.5%;  Score 462.6;  DB 14;  Length 178887;
  Best Local Similarity   62.1%;  Pred. No. 7e-83;
  Matches  890;  Conservative    1;  Mismatches  485;  Indels   57;  Gaps    8;

Qy        600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
              || || |||| |||||||||||        | | ||  |||||||| |||||||  |||
Db      90620 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 90561

Qy        660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
              ||||||  | | |   | ||||||||  |||  ||||||| |||||||  |||  || ||
Db      90560 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 90501

Qy        720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
              |||||| ||||||||||||||||||||  | |  || ||||| || | |||||||
Db      90500 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 90441

Qy        780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
              |  |||| |||||||||  |||| || | | |||| ||| |||   | | |||||||||
Db      90440 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 90383

Qy        840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
              |||| | |||| |||  ||  || |||||||||| |||||||||| ||||||||||
Db      90382 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 90324
```

```
Qy        900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
Db      90323 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 90264

Qy        960 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
Db      90263 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTGTGGGACTGCCACTGGAGTGCAGTG 90205

Qy       1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
Db      90204 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 90145

Qy       1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
Db      90144 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 90085

Qy       1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
Db      90084 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 90025

Qy       1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
Db      90024 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 89965

Qy       1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
Db      89964 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 89913

Qy       1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
Db      89912 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 89867

Qy       1380 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
Db      89866 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 89828

Qy       1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
Db      89827 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 89768

Qy       1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
Db      89767 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 89708

Qy       1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
Db      89707 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 89653

Qy       1620 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
Db      89652 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 89593

Qy       1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
Db      89592 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 89533

Qy       1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
Db      89532 TTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 89475

Qy       1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
Db      89474 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 89415

Qy       1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
Db      89414 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 89355

Qy       1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
Db      89354 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACGCCACTATACTC 89295
```

```
Qy       1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 2029
Db      89294 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAAACA 89242
```

```
RESULT 7
AC009244
LOCUS       AC009244               190162 bp    DNA     linear   PRI 26-MAY-2002
DEFINITION  Homo sapiens chromosome 7 clone RP11-305M3, complete sequence.
ACCESSION   AC009244
VERSION     AC009244.25  GI:21217399
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190162)
  AUTHORS   Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-1999) Human Genome Center, University of
            Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
            Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   4  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 26, 2002 this sequence version replaced gi:11038533.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-7
                Center clone name: RP11-305M3 (djs195)
            ------------- Summary Statistics
                Assembly program: Phrap; version 0.990319
                Consensus quality: 189969 bases at least Q40
                Consensus quality: 190110 bases at least Q30
                Consensus quality: 190161 bases at least Q20
                Insert size: 190162; sum-of-contigs
                Quality coverage: 9.7x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5':  RP11-382M23 (UWGC:djs734g) AC093149 1340-bp overlap
                3':  RP11-448A19 (UWGC:djs705) AC078846 8765-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
```

all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
21251	20929	12417	12174	8696	8794
12148	11918	2067	2041	6	<800
199	<800	5666	5826	1775	1773
3275	3221	254	<800	2000	1991
20379	20929	3818	3836	6313	6324
15555	15150	747	735	3256	3226
675	<800	9196	9237	325	<800
5493	5475	4154	4101	12989	12723
516	<800	5820	5826	5955	5951
5810	5831	118	<800	611	<800
9488	9435	2524	2359	1995	1991
2908	2935	8844	8768	11945	12009
779	840	332	<800	5577	5593
821	840	564	<800	4139	4124
315	<800	26	<800	955	940
3640	3671	4675	4644	2961	2940
428	<800	1801	1803	5264	5205
1144	1119	117	<800	1535	1534
4684	4717	6797	6799	3313	3383
20	<800	3146	3177	3760	3761
492	<800	2376	2359	3061	3110
10935	10879	210	<800	9605	9505
5921	6010	5135	5110	335	<800
529	<800	744	735	2190	2194
4354	4384	881	947	5117	5205
1311	1293	413	<800	13558	13434
10	<800	6389	6456	9367	9505
374	<800	296	<800	8080	8019
1497	1480	7160	7119	4276	4243
4079	4083	4846	4822	7830	8019
337	<800	2914	2924	231	<800
4483	4384	3803	3836	533	<800
1530	1480	1832	1803	668	<800
249	<800	674	<800	12314	12009
8831	8995	2515	2513	18561	18820
244	<800	2163	2137	11006	11013
1019	1000	466	<800	8756	8794
841	840	927	947	7	<800
8114	8086	5963	5826		
9184	9435	7440	7660		
7121	7263	5159	4822		
1122	1119	677	<800		
49	<800	1399	1373		
11928	11918	1565	1554		
1499	1480	2388	2513		
2585	2575	4826	4822		
699	<800	7661	7660		
		3624	3705		
		8693	8768		
		1749	1803		
		1642	1635		
		5295	5362		
		10628	10715		
		3227	3177		
		589	<800		

Query Match 18.5%; Score 462.6; DB 8; Length 190162;
Best Local Similarity 61.3%; Pred. No. 6.9e-83;
Matches 900; Conservative 1; Mismatches 520; Indels 46; Gaps 8;

```
Qy      604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
            ||||  ||| ||       |||| || | ||||||||  ||||||||||  | |  ||||
Db    31687 TTTTCCTTTTTTTAAAAGTAGAAACGGAGTCTTGCTGGGTCACCCAGGCTGGGGTGCAGT 31746

Qy      664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
            |||||  | |  |||||||| ||| |  ||||  |  || | ||||  |  || |||
Db    31747 AGTGCAGTCGTGTTTCACTGCATCCTTCAATTCCTAGGAACAGGTGATCCTCCCACTTCA 31806
```

```
Qy       724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db     31807 GTCTTCTGAGTAGCTGGGACTTCAGGCATGTGCCACCATGCCCAGCTAATTTAAAAAAAA 31866

Qy       784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db     31867 TGTTTGAAGACACAGGGTCTTGCTGTGTTGCTCAGGCTGGTCTCAAACAGCCAGCCTCAA 31926

Qy       844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db     31927 GTGATCCTCCTGCCTCATCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCC 31986

Qy       904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db     31987 AGCCCATTTCTTTTTTCTGTTTTTTTTTTAAATTTAATTTAATTTTTATTTTTTATTTT 32046

Qy       964 TTACTGCCTATTTTCTTCCTT-TCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db     32047 TATTTTTATTTTTATTTATTTATTTTTTTTGAGATGAAGTCTCACTCTGTCGCCCAGGCT 32106

Qy      1023 TCCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 1068
Db     32107 GGAGTGCAGTGGCATGATCTCGGCTCACTGAAACCTCTGCCTCCCAGGTTCAAGCGATTC 32166

Qy      1069 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAAT 1128
Db     32167 TCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCCTCCACTATACCCGGCTAA- 32225

Qy      1129 TTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCC 1188
Db     32226 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGTCAGGCTGGTCTCGAACTCC 32285

Qy      1189 TGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTC 1248
Db     32286 TGACCTCAGATGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCC 32345

Qy      1249 ACTGTGCCTGGCCGATTACTGTCTATTTT---CTTTATTGCTATATCCCCAGATCTAGAG 1305
Db     32346 ACCGCACCCGGCTGCCCCATTTTTTTTTTTAAACAAATGCTTACTGACCTACGCTAAGAG 32405

Qy      1306 CAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATA 1365
Db     32406 CTGTAAGGGCAAATGAGACAATGTCCTTACCCTCAGAATCTTATATTTTGGTGAGATAAT 32465

Qy      1366 AACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTG 1425
Db     32466 ACACATATATAAATAATGATGATACAAGGACATATATAATAAATGCCATGACAGATTATA 32525

Qy      1426 CATTCTGCTTTTTCAC-TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1484
Db     32526 AATCCATATTCACTACACAAACAAGAGGCTTAGGAAAGTTCAGATTCCCACTTTGGGAAG 32585

Qy      1485 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1544
Db     32586 CTGAGGCAGGTGGATCACCTGAGGTCAGGAATTCA--AGACCAGCGTGGCCAACATGGTG 32643

Qy      1545 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1604
Db     32644 AAACCCTGTCTTTACTAAGAGTACAAAAATTAGCTGAGTGTGGTGGTGCAGGCCTGTAAT 32703

Qy      1605 CCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTT 1664
Db     32704 CCTAGCTACTCATGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGCTGCA 32763

Qy      1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTG---------------------- 1702
Db     32764 GTGAGCCGAGATCGCGCCACTGCACTCTAGCCTGGGTGACAGAGTGAGACTCCATCTCAA 32823

Qy      1703 TTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTAC 1762
Db     32824 AAAAAAAAAAAAAAAAAAAAAGAGGCCAGGCGCGGTGGCTCACACCTGTAATCCCAGCAC 32883

Qy      1763 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAA 1822
Db     32884 TTTGGGAGCCTGAGGTGGGCGGATCA--TGAGGTCAGAAGTTCGAGACCAGCCTGGCCAA 32941

Qy      1823 TATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGC 1882
Db     32942 TATGGTGAAACCCCGTCTTTACTAAAAGTACAAAAATTAGCTGGGCGTGGTGCCACGCAC 33001

Qy      1883 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 1942
Db     33002 CTGTAGTCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGA 33061

Qy      1943 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCT 2002
Db     33062 GGTTGCAGTGAGCCGAGATTATGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCC 33121

Qy      2003 ATCTCAAAAAAATAAATAAATAAAATA 2029
Db     33122 GTCTCAAAAAAAAAAAAGAAAAGAAAA 33148
```

RESULT 8
AP000589

```
LOCUS       AP000589              123192 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-105N5 map 11q13, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP000589
VERSION     AP000589.3  GI:8118795
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 123,192 genomic DNA of 11q13
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997479.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-105N5
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 115116 bases at least Q40
                Consensus quality: 118340 bases at least Q30
                Consensus quality: 120191 bases at least Q20
                Insert size: 121392; sum-of-contigs
                Quality coverage: 5.08x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
```

```
as soon as it is available and the accession number will be
preserved
         1     15827 contig of  15827 bp in length
     15928     30484 contig of  14557 bp in length
     30585     40602 contig of  10018 bp in length
     40703     51912 contig of  11210 bp in length
     52013     60022 contig of   8010 bp in length
     60123     70438 contig of  10316 bp in length
     70539     77945 contig of   7407 bp in length
     78046     85717 contig of   7672 bp in length
     85818     92399 contig of   6582 bp in length
     92500     98215 contig of   5716 bp in length
     98316    104148 contig of   5833 bp in length
    104249    109166 contig of   4918 bp in length
    109267    112457 contig of   3191 bp in length
    112558    115076 contig of   2519 bp in length
    115177    117063 contig of   1887 bp in length
    117164    118652 contig of   1489 bp in length
    118753    120425 contig of   1673 bp in length
    120526    121843 contig of   1318 bp in length
    121944    123192 contig of   1249 bp in length
Sequence updated (18-Oct-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    15827: contig of 15827 bp in length
*    15828    15927: gap of 100 bp
*    15928    30484: contig of 14557 bp in length
*    30485    30584: gap of 100 bp
*    30585    40602: contig of 10018 bp in length
*    40603    40702: gap of 100 bp
*    40703    51912: contig of 11210 bp in length
*    51913    52012: gap of 100 bp
*    52013    60022: contig of 8010 bp in length
*    60023    60122: gap of 100 bp
*    60123    70438: contig of 10316 bp in length
*    70439    70538: gap of 100 bp
*    70539    77945: contig of 7407 bp in length
*    77946    78045: gap of 100 bp
*    78046    85717: contig of 7672 bp in length
*    85718    85817: gap of 100 bp
*    85818    92399: contig of 6582 bp in length
*    92400    92499: gap of 100 bp
*    92500    98215: contig of 5716 bp in length
*    98216    98315: gap of 100 bp
*    98316   104148: contig of 5833 bp in length
*   104149   104248: gap of 100 bp
*   104249   109166: contig of 4918 bp in length
*   109167   109266: gap of 100 bp
*   109267   112457: contig of 3191 bp in length
*   112458   112557: gap of 100 bp
*   112558   115076: contig of 2519 bp in length
*   115077   115176: gap of 100 bp
*   115177   117063: contig of 1887 bp in length
*   117064   117163: gap of 100 bp
*   117164   118652: contig of 1489 bp in length
*   118653   118752: gap of 100 bp
*   118753   120425: contig of 1673 bp in length
*   120426   120525: gap of 100 bp
*   120526   121843: contig of 1318 bp in length
*   121844   121943: gap of 100 bp
*   121944   123192: contig of 1249 bp in length.
FEATURES             Location/Qualifiers
     source          1. .123192
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q13"
                     /clone="CMB9-105N5"
     misc_feature    1. .15827
                     /note="assembly_fragment"
     misc_feature    15928. .30484
                     /note="assembly_fragment"
     misc_feature    30585. .40602
                     /note="assembly_fragment"
     misc_feature    40703. .51912
                     /note="assembly_fragment"
     misc_feature    52013. .60022
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    60123. .70438
                     /note="assembly_fragment"
     misc_feature    70539. .77945
                     /note="assembly_fragment"
     misc_feature    78046. .85717
                     /note="assembly_fragment"
     misc_feature    85818. .92399
                     /note="assembly_fragment"
     misc_feature    92500. .98215
                     /note="assembly_fragment"
     misc_feature    98316. .104148
                     /note="assembly_fragment"
     misc_feature    104249. .109166
                     /note="assembly_fragment"
     misc_feature    109267. .112457
                     /note="assembly_fragment"
     misc_feature    112558. .115076
                     /note="assembly_fragment"
     misc_feature    115177. .117063
                     /note="assembly_fragment"
     misc_feature    117164. .118652
                     /note="assembly_fragment"
     misc_feature    118753. .120425
                     /note="assembly_fragment"
     misc_feature    120526. .121843
                     /note="assembly_fragment"
     misc_feature    121944. .123192
                     /note="assembly_fragment"
ORIGIN

  Query Match            18.4%;  Score 460.6;  DB 14;  Length 123192;
  Best Local Similarity  59.8%;  Pred. No. 1.9e-82;
  Matches  882;  Conservative    1;  Mismatches  575;  Indels   17;  Gaps    6;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 656
              |||   ||||||||| || || ||| ||||||| || |  ||| |||  |||||||  ||
Db      63120 TTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 63179

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
              |  || |||  ||| |   | ||||  || ||||  ||||| | | |||||||  |   |
Db      63180 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 63239

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
              |||||||||||| |||||||| || |||  || ||| |||||||||  ||| |||
Db      63240 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 63299

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
              |   || || |||||||| |||| |  | | ||||| |||||| ||||  | ||||||
Db      63300 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 63356

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
              | ||| |  |||| || |||| ||| |||||||||||||||||| ||| ||||||||
Db      63357 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 63416

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
              || |  |||| |  || |  |||  |     ||  ||   || |          |
Db      63417 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 63476

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
              |      ||     || |||| ||       | ||           |   ||  |   |
```

```
Db      63477 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 63536

Qy       1017 TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 1076

Db      63537 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCT 63596

Qy       1077 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--G 1134

Db      63597 CAGCCTCCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGT 63656

Qy       1135 TATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCT 1194

Db      63657 ATTTTTTAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 63716

Qy       1195 CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTG 1254

Db      63717 CAGGTGATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCG 63776

Qy       1255 CCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314

Db      63777 CCCGATGCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TA 63835

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374

Db      63836 TCATTTCCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATG 63895

Qy       1375 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434

Db      63896 ATTATTATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATT 63955

Qy       1435 TTTTCA---------CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTAT 1485

Db      63956 ATATCAATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAA 64015

Qy       1486 TGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTA 1545

Db      64016 TGGGTGTTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGT 64075

Qy       1546 TACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACC 1605

Db      64076 TTCTAGAACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGT 64135

Qy       1606 CTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGT-TTTTTTTTAATGTT 1664

Db      64136 CCATGGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGC 64195

Qy       1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724

Db      64196 CAAAAATGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGA 64255

Qy       1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784

Db      64256 AGGCCAGGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGG 64315

Qy       1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844

Db      64316 ATTACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTAC 64375

Qy       1845 TAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 1904

Db      64376 TAAAAATACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGA 64435

Qy       1905 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 1964

Db      64436 GGCTGAGGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGT 64495

Qy       1965 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATA 2024

Db      64496 GTCACTGCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATA 64555

Qy       2025 AAATAAAGGATCGGAGAGAAACAAAACTAATAAGA 2059

Db      64556 AAAATAATGAAAATAATAAAATGATGTTAAAGATA 64590
```

```
RESULT 9
AC022408
LOCUS       AC022408              150266 bp    DNA     linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-118J16 map 15q15, complete
            sequence.
ACCESSION   AC022408
VERSION     AC022408.6  GI:15217198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Aug 21, 2001 this sequence version replaced gi:11225361.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Note:data from AC021753 [Drafting center UWMSC] and AC012652
            [Drafting center UWMSC] and 580I7 [Drafting center WUGSC] were
            added for finishing.
FEATURES             Location/Qualifiers
     source          1. .150266
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-118J16"
                     /clone_lib="RPCI human BAC library 11"
                     /note="This clone overlaps RP11-46M12 AC012652 and
                     RP11-129I12 AC021753 and RP11-580I7 AC068981.Data from
                     overlapping BACs were combined and the consensus sequence
                     determined from RP11-118J16."
     misc_feature    1. .150266
                     /note="overlap with RP11-580I7 AC068981."
     misc_feature    1. .100397
                     /note="overlap with RP11-129I12 AC021753."
     unsure          49355. .49359
                     /note="low quality data."
     misc_feature    100628. .150266
                     /note="overlap with RP11-46M12 AC012652."
ORIGIN
```

Query Match 18.4%; Score 460.4; DB 8; Length 150266;
Best Local Similarity 62.0%; Pred. No. 2e-82;
Matches 889; Conservative 1; Mismatches 487; Indels 57; Gaps 8;

```
Qy       600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
Db     52138 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 52197

Qy       660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
Db     52198 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 52257

Qy       720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
Db     52258 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 52317

Qy       780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
Db     52318 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 52375

Qy       840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
Db     52376 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCAA 52434

Qy       900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
Db     52435 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 52494

Qy       960 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
Db     52495 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 52553

Qy      1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
Db     52554 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 52613

Qy      1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
Db     52614 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 52673

Qy      1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
Db     52674 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 52733

Qy      1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
Db     52734 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 52793

Qy      1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
Db     52794 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 52845

Qy      1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
Db     52846 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 52891

Qy      1380 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
Db     52892 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 52930

Qy      1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
Db     52931 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 52990

Qy      1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
Db     52991 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 53050

Qy      1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
Db     53051 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 53105

Qy      1620 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
Db     53106 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 53165

Qy      1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
Db     53166 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 53225

Qy      1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
Db     53226 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 53283

Qy      1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
Db     53284 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 53343

Qy      1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
Db     53344 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 53403

Qy      1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
Db     53404 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACACCACTATACTC 53463

Qy      1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db     53464 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAACATTCA 53517
```

RESULT 10
AC021753
LOCUS AC021753 167996 bp DNA linear PRI 29-OCT-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-129I12 map 15q15, complete sequence.
ACCESSION AC021753
VERSION AC021753.7 GI:16507213
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167996)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167996)
AUTHORS Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 167996)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
COMMENT On Oct 29, 2001 this sequence version replaced gi:11136698.
---------- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
---------- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads

```
            Chemistry: Dye-primer Big Dye; 10% of reads
            Assembly program: Phrap; version 0.990399
            Note: data from AC022408 [Drafting center UWMSC] and AC020661
          were added for fimishing.
FEATURES             Location/Qualifiers
     source          1. .167996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-129I12"
                     /clone_lib="RPCI Human BAC library 11"
                     /note="This clone overlaps RP11-118J16 AC022408 and
                     RP11-540O11 AC020661. Data from overlapping BACs were
                     combined and the consensus sequence determined from
                     RP11-129I12 to the extent possible."
     misc_feature    1. .63402
                     /note="overlap with RP11-540O11 AC020661."
     misc_feature    67596. .167996
                     /note="overlap with RP11-118J16 AC022408."
ORIGIN

  Query Match             18.4%;  Score 459.4;  DB 8;  Length 167996;
  Best Local Similarity   62.0%;  Pred. No. 3.1e-82;
  Matches  888;  Conservative    1;  Mismatches  487;  Indels   57;  Gaps    8;

Qy        600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
              || || |||| ||||||||||        |  | ||  |||||||| |||||||  |||
Db     119755 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 119814

Qy        660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
              ||||||  | | |   | ||||||||  |||  |||||||  |||||||  |||  || ||
Db     119815 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 119874

Qy        720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
              |||||| |||||||||||||||||||||   | |   || |||||  || | |||||||
Db     119875 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 119934

Qy        780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
              |  |||| |||||||||  |||| || | |  |||| ||| |||   |.| |||||||||
Db     119935 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 119992

Qy        840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
              |||| | |||| |||   ||  || |||||||||||  ||||||||||  ||||||||||
Db     119993 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 120051

Qy        900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
              |      |   ||   | |  |  |       |  |  ||| |||      | |||
Db     120052 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 120111

Qy        960 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
                 ||  ||  | |||    |     |||  |       |     ||||  |   |
Db     120112 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 120170

Qy       1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
                |     || |||||||||||||| ||||||||||||||:|||||||||||||  ||
Db     120171 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 120230

Qy       1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
              |||||||||||||||| |||||| | | || |||||| ||   ||||||||||||||||
Db     120231 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 120290

Qy       1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
              || | ||||   | |||| |||||||||| |||||||||| || ||||||||||||||  ||
Db     120291 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 120350

Qy       1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
              ||||  |||||||||||||||||||||| || |||| |||||||| ||| |  || ||
Db     120351 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 120410

Qy       1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
              || |  |   | |||   | |  |||   || ||| |    |  ||| |

Db     120411 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 120462

Qy       1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
                          |  |  ||  |  |  ||  | || | | |  ||      ||
Db     120463 -------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 120508

Qy       1380 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
              |    |   |||||| |  |    | |      ||                      |
Db     120509 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 120547

Qy       1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
              | | |  |||  | |  |  |  |  ||  | ||  |  |  ||  |  |  |
Db     120548 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 120607

Qy       1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
              |  |     | ||    |  || ||| | ||  |    ||    ||  ||   |  |  |
Db     120608 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 120667

Qy       1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
                | | |     || ||| | ||||  ||   |||  |  ||  |     |  | |
Db     120668 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 120722

Qy       1620 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
                |  ||| ||   || |     |   |||| |  | ||       |    | || |||
Db     120723 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 120782

Qy       1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
              ||  | | |   | | | ||||    |  ||       ||| | || ||||  ||||||
Db     120783 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 120842

Qy       1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
              ||| |||| ||||||||  | ||| |||||||  || |||||||  ||||||||||  |||||
Db     120843 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 120900

Qy       1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
              ||  | ||||||||||||||  ||| ||||||||||||  ||||||||||||||||||||
Db     120901 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 120960

Qy       1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
              |||||||||||||| ||| |  |||||| | |||||||||| ||||||||||||| |
Db     120961 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 121020

Qy       1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
              |||||||||||||||| |||||| || | ||||||||||||||||| | |||||  ||||
Db     121021 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACGCCACTATACTC 121080

Qy       1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 2029
              |||||||||||||| ||| ||||||  |||||||||  ||| ||| |||| |
Db     121081 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAAACA 121133


RESULT 11
AC003689
LOCUS       AC003689              137693 bp    DNA     linear   PRI 21-NOV-1998
DEFINITION  Homo sapiens Chromosome 11q12.2 PAC clone pDJ1081b4 containing
            human mRNA for T-cell glycoprotein CD6, complete sequence.
ACCESSION   AC003689
VERSION     AC003689.1  GI:3900834
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 137693)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
```

REFERENCE 2 (bases 1 to 137693)
AUTHORS Evans,G.A., Athanasiou,M., Basit,M., Bradbury,P., Brignac,S., Bumesiter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwester Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE 3 (bases 1 to 137693)
AUTHORS Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Butler,C., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT On Nov 21, 1998 this sequence version replaced gi:2695567.
Further information regarding the map of this region or annotation of pDJ1081b4 can be found at http://gestec.swmed.edu/chromosol.htm
IMPORTANT: This submission contains the entire insert of clone pDJ1081b4. pDJ1081b4 comes from the RPCI-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.98. In addition, attempts have been made to assure over 99% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.
CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome 11p12.2 Best's disease region mapped between STS D11S461 and EST AHNAK. This region spans over 1.5 Mbp.
MARKER CONFIRMATION: ESTs/STSs sequence confirmed; CD6, D11S461
MAPPED CLONE OVERLAP: PACs pDJ606g6 and pDJ59L21.

FEATURES Location/Qualifiers

source	1. .137693
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
repeat_region	164. .454
	/rpt_family="Alu"
repeat_region	complement(1543. .1835)
	/rpt_family="Alu"
repeat_region	complement(2058. .2360)
	/rpt_family="Alu"
repeat_region	complement(3515. .3796)
	/rpt_family="Alu"
repeat_region	complement(5954. .6244)
	/rpt_family="Alu"
repeat_region	complement(6962. .7215)
	/rpt_family="Alu"
repeat_region	complement(8088. .8263)
	/rpt_family="Alu"
repeat_region	complement(12551. .12719)
	/rpt_family="Alu"
repeat_region	13480. .13676
	/rpt_family="MER33"
repeat_region	13677. .13985
	/rpt_family="Alu"
repeat_region	complement(15815. .15950)
	/rpt_family="Alu"
repeat_region	complement(17641. .17934)
	/rpt_family="Alu"
repeat_region	18423. .18693
	/rpt_family="Alu"
repeat_region	18722. .19008
	/rpt_family="Alu"
repeat_region	complement(21315. .21594)
	/rpt_family="Alu"
repeat_region	complement(21636. .22481)
	/rpt_family="MER42"
repeat_region	complement(23396. .23697)
	/rpt_family="Alu"
repeat_region	25083. .25349
	/rpt_family="Alu"
repeat_region	complement(28459. .28738)
	/rpt_family="Alu"
repeat_region	29031. .29103
	/rpt_family="Alu"
repeat_region	complement(29491. .29785)
	/rpt_family="Alu"
repeat_region	complement(30052. .30172)
	/rpt_family="Alu"
repeat_region	complement(30261. .30640)
	/rpt_family="Alu"
repeat_region	30885. .31193
	/rpt_family="Alu"
repeat_region	32498. .32814
	/rpt_family="Alu"
repeat_region	complement(33478. .33782)
	/rpt_family="Alu"
repeat_region	34395. .34556
	/rpt_family="MIR"
repeat_region	35452. .35745
	/rpt_family="Alu"
repeat_region	complement(35739. .35983)
	/rpt_family="MER33"
repeat_region	complement(36141. .36244)
	/rpt_family="MER3"
repeat_region	complement(36206. .36244)
	/rpt_family="MER33"
repeat_region	37244. .37491
	/rpt_family="Alu"
repeat_region	38123. .38398
	/rpt_family="Alu"
repeat_region	complement(39547. .39837)
	/rpt_family="Alu"
repeat_region	40124. .40270
	/rpt_family="MER45"
repeat_region	40625. .40913
	/rpt_family="Alu"
repeat_region	complement(41015. .41075)
	/rpt_family="MLT1"
repeat_region	complement(41333. .41615)
	/rpt_family="Alu"
repeat_region	41812. .42105
	/rpt_family="Alu"
repeat_region	44026. .44309
	/rpt_family="Alu"
repeat_region	complement(47659. .47752)
	/rpt_family="MER5"
repeat_region	50186. .50330
	/rpt_family="MIR"
repeat_region	52843. .53128
	/rpt_family="Alu"
repeat_region	53437. .53590
	/rpt_family="MIR"
repeat_region	53688. .53967
	/rpt_family="Alu"
repeat_region	56946. .57144
	/rpt_family="Alu"
repeat_region	59439. .59731
	/rpt_family="Alu"
repeat_region	complement(61583. .61649)
	/rpt_family="MIR"
repeat_region	complement(62025. .62159)
	/rpt_family="MIR"
repeat_region	complement(62743. .63037)
	/rpt_family="Alu"

```
     repeat_region   complement(63993. .64088)
                     /rpt_family="MIR"
     repeat_region   complement(66479. .66870)
                     /rpt_family="Alu"
     repeat_region   67416. .67584
                     /rpt_family="MER42"
     repeat_region   67594. .67887
                     /rpt_family="Alu"
     repeat_region   68119. .68182
                     /rpt_family="MER42"
     repeat_region   complement(69234. .69519)
                     /rpt_family="Alu"
     repeat_region   complement(69560. .69821)
                     /rpt_family="Alu"
     repeat_region   complement(70420. .70676)
                     /rpt_family="Alu"
     repeat_region   complement(70696. .70906)
                     /rpt_family="MER42"
     repeat_region   71247. .71537
                     /rpt_family="Alu"
     repeat_region   complement(71520. .71655)
                     /rpt_family="MER42"
     repeat_region   72778. .73053
                     /rpt_family="Alu"
     repeat_region   complement(75074. .75330)
                     /rpt_family="Alu"
     repeat_region   complement(76584. .76879)
                     /rpt_family="Alu"
     repeat_region   complement(76944. .77232)
                     /rpt_family="Alu"
     repeat_region   77719. .77994
                     /rpt_family="Alu"
     repeat_region   78902. .79179
                     /rpt_family="Alu"

  Query Match             18.3%;  Score 457.4;  DB 8;  Length 137693;
  Best Local Similarity   59.7%;  Pred. No. 8e-82;
  Matches  880;  Conservative    1;  Mismatches  577;  Indels   17;  Gaps    6;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 656
              |||   | ||| ||| || || ||| |||||||  || |  ||| |||  |||||||  ||
Db      76572 TTTCTTTCTTTCTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 76631

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
              |  || |||  ||| |   | ||||  || ||||  ||||| | | |||||||  |   |
Db      76632 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 76691

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
              |||||||||||| ||||||| || |||  || ||| ||||||||||  ||| |||
Db      76692 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 76751

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
              |   || || |||||||| |||| |  | | |||||  |||||| ||||  | |||||||
Db      76752 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 76808

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
              | ||| |  |||| || |||| ||| ||||||||||||||||||| ||| ||||||||
Db      76809 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 76868

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
              || |  ||||  |  || |  |||  |      ||   ||     || |          |
Db      76869 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 76928

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
              |      ||     || |||| ||       | ||           |   ||  |   |
Db      76929 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 76988

Qy       1017 TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 1076
                |     || |||||||||| |||||||||  |||||| |:|||||||||||||
Db      76989 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCT 77048

Qy       1077 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--G 1134
              ||||||| |||||||||| |||||| | | || |||||| |  |||||| ||||||
```

```
Db      77049 CAGCCTCCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGT 77108

Qy       1135 TATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCT 1194
              |||||||  |||||  ||||||  |||||||||  ||||||||||| ||||||||||||
Db      77109 ATTTTTTAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 77168

Qy       1195 CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTG 1254
              || ||||||  |||||||||||  || ||||||||||||  ||||  ||||  ||||  ||| | |
Db      77169 CAGGTGATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCG 77228

Qy       1255 CCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314
              || |      |   |   ||     |    |||||      |  |     |  || |
Db      77229 CCCGATGCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TA 77287

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374
              |||  |  |  |  |  ||           |    | |          || |   | || |
Db      77288 TCATTTCCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATG 77347

Qy       1375 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434
              || || ||   | | ||     |  | | ||   | || | |   |  ||| | | |
Db      77348 ATTATTATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATT 77407

Qy       1435 TTTTCA---------CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTAT 1485
              | |||           | || | |  |  |  |||   |  |||||        | |
Db      77408 ATATCAATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAA 77467

Qy       1486 TGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTA 1545
              ||    | |  ||  |   |  | ||  | ||  |  | |       |||   |
Db      77468 TGGGTGTTTAGGTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGT 77527

Qy       1546 TACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACC 1605
              | |      | |      |||| |  |    || |    | |     |  | |   |
Db      77528 TTCTAGAACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGT 77587

Qy       1606 CTCTCCCCAGAGAAATACACACACACACACACTCACACACAGT-TTTTTTTTAATGTT 1664
              | |     ||||  |    | | |  |  | |  ||  | |     ||   |
Db      77588 CCATGGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGC 77647

Qy       1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724
                || |   ||||   |          |   | ||| ||  | || |  |  |
Db      77648 CAAAAATGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGA 77707

Qy       1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784
              |   ||    ||| |||| |  ||||| |  || ||||||||| || |||  |||||
Db      77708 AGGCCAGGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGG 77767

Qy       1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844
              || ||||||||| || |||||||||||||||||| ||| |||||||||| || ||||||||
Db      77768 ATTACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTAC 77827

Qy       1845 TAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 1904
              |||||||||||||||||||||| || ||||||| ||||||| |||||||||| ||||
Db      77828 TAAAAATACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGA 77887

Qy       1905 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 1964
              |||||||||  ||||| |||||||| |||||||  ||||||||||||||||||||||||
Db      77888 GGCTGAGGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGT 77947

Qy       1965 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATA 2024
              |||||||||||||||||| ||| | | ||| ||| |||||||| ||  ||||||||
Db      77948 GTCACTGCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATA 78007

Qy       2025 AAATAAAGGATCGGAGAGAAACAAAACTAATAAGA 2059
              |||  || ||    |  |||  |   |||  | |
Db      78008 AAAATAATGAAAATAATAAAATGATGTTAAAGATA 78042


RESULT 12
AP003721
LOCUS       AP003721              201460 bp    DNA     linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-881M11, complete
            sequence.
ACCESSION   AP003721
```

```
VERSION     AP003721.3  GI:31790667
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 201460)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 16, 2003 this sequence version replaced gi:28189463.
FEATURES             Location/Qualifiers
     source          1. .201460
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-881M11"
ORIGIN

  Query Match           18.3%;  Score 456.2;  DB 8;  Length 201460;
  Best Local Similarity  59.7%;  Pred. No. 1.3e-81;
  Matches  877;  Conservative    1;  Mismatches  574;  Indels   17;  Gaps    6;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAG 662
Db     144460 TTTCTTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGAGTGCAA 144519

Qy        663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 722
Db     144520 TGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCCTGCCTC 144579

Qy        723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db     144580 AGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTTGTA--- 144636

Qy        783 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db     144637 TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTGACCTCA 144696

Qy        843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db     144697 GGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCATCACGCC 144756

Qy        903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db     144757 CGGCCTCACATATTCTTATCCTAATATGTTTTCACGCTTTGAAGTCCTACAGTAATCATC 144816

Qy        963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db     144817 CTTTGATTCTTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCAATGGCG 144876

Qy       1023 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 1082
Db     144877 CAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCTCAGCCT 144936

Qy       1083 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--GTATTTT 1140
Db     144937 CCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGTATTTTT 144996

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
```

```
Db     144997 TAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTG 145056

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db     145057 ATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGAT 145116

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db     145117 GCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TATCATTT 145175

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db     145176 CCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATGATTATT 145235

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db     145236 ATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATTATATCA 145295

Qy       1441 C---------TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTAC 1491
Db     145296 ATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAATGGGTG 145355

Qy       1492 TGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATC 1551
Db     145356 TTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGTTTCTAG 145415

Qy       1552 CTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCC 1611
Db     145416 AACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGTCCATGG 145475

Qy       1612 CCAGAGAAATACACACACACACACACACTCACACACAGT-TTTTTTTTAATGTTTGCAAC 1670
Db     145476 AGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGCCAAAAA 145535

Qy       1671 TAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTG 1730
Db     145536 TGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGAAGGCCA 145595

Qy       1731 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 1790
Db     145596 GGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGGATTACT 145655

Qy       1791 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 1850
Db     145656 TGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTACTAAAAA 145715

Qy       1851 TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGA 1910
Db     145716 TACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGAGGCTGA 145775

Qy       1911 GGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACT 1970
Db     145776 GGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGTCACT 145835

Qy       1971 GCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db     145836 GCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATAAAAATA 145895

Qy       2031 AGGATCGGAGAGAAACAAAACTAATAAGA 2059
Db     145896 ATGAAAATAATAAAATGATGTTAAAGATA 145924
```

```
RESULT 13
AC116933
LOCUS       AC116933              195986 bp    DNA     linear   HTG 21-DEC-2002
DEFINITION  Papio anubis clone RP41-1C24, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC116933
VERSION     AC116933.2  GI:27356688
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
```

```
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 195986)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Dec 21, 2002 this sequence version replaced gi:19909407.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cqw
                 Center clone name: 001C24

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 195122 bases at least Q40
                 Consensus quality: 195370 bases at least Q30
                 Consensus quality: 195594 bases at least Q20
                 Insert size: 186000; agarose-fp
                 Insert size: 195686; sum-of-contigs
                 Quality coverage: 9.19x in Q20 bases; agarose-fp
                 Quality coverage: 8.74x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    76152: contig of 76152 bp in length
            *    76153    76252: gap of unknown length
            *    76253   126668: contig of 50416 bp in length
            *   126669   126768: gap of unknown length
            *   126769   177394: contig of 50626 bp in length
            *   177395   177494: gap of unknown length
            *   177495   195986: contig of 18492 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195986
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-1C24"
                     /clone_lib="RP41"
     misc_feature    1. .76152
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .54626
                     /note="clone overlaps with GenBank Accession Number
                     AC116934 clone RP41-206F5 (center project name cqv)"
     gap             76153. .76252
                     /estimated_length=unknown
     misc_feature    76253. .126668
                     /note="assembly_fragment"
     gap             126669. .126768
                     /estimated_length=unknown
     misc_feature    126769. .177394
                     /note="assembly_fragment"
     gap             177395. .177494
                     /estimated_length=unknown
     misc_feature    177495. .195986
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match            18.1%;  Score 451.2;  DB 14;  Length 195986;
  Best Local Similarity  61.1%;  Pred. No. 1.3e-80;
  Matches  891;  Conservative    1;  Mismatches  514;  Indels   52;  Gaps    8;

Qy       604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
Db    180502 TTTTCCTTTTTTTTTAAGTAGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGT 180561

Qy       664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
Db    180562 AGTGCAATAGTGTTTCCCTGCATCCTTCAATTCCTGGGAACAGGTGATCCTCCCACCTCA 180621

Qy       724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db    180622 GTCTTCTGAGTAGCTGGGACTTTAGGCACATGCCACCATGCCCAGCTAACTTAAAAAATT 180681

Qy       784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db    180682 TTTTTGAAGACACAGGGTCTTGCTCTGTTGCTCAGGCTACTCTCAAACAGCCGGCCTCAA 180741

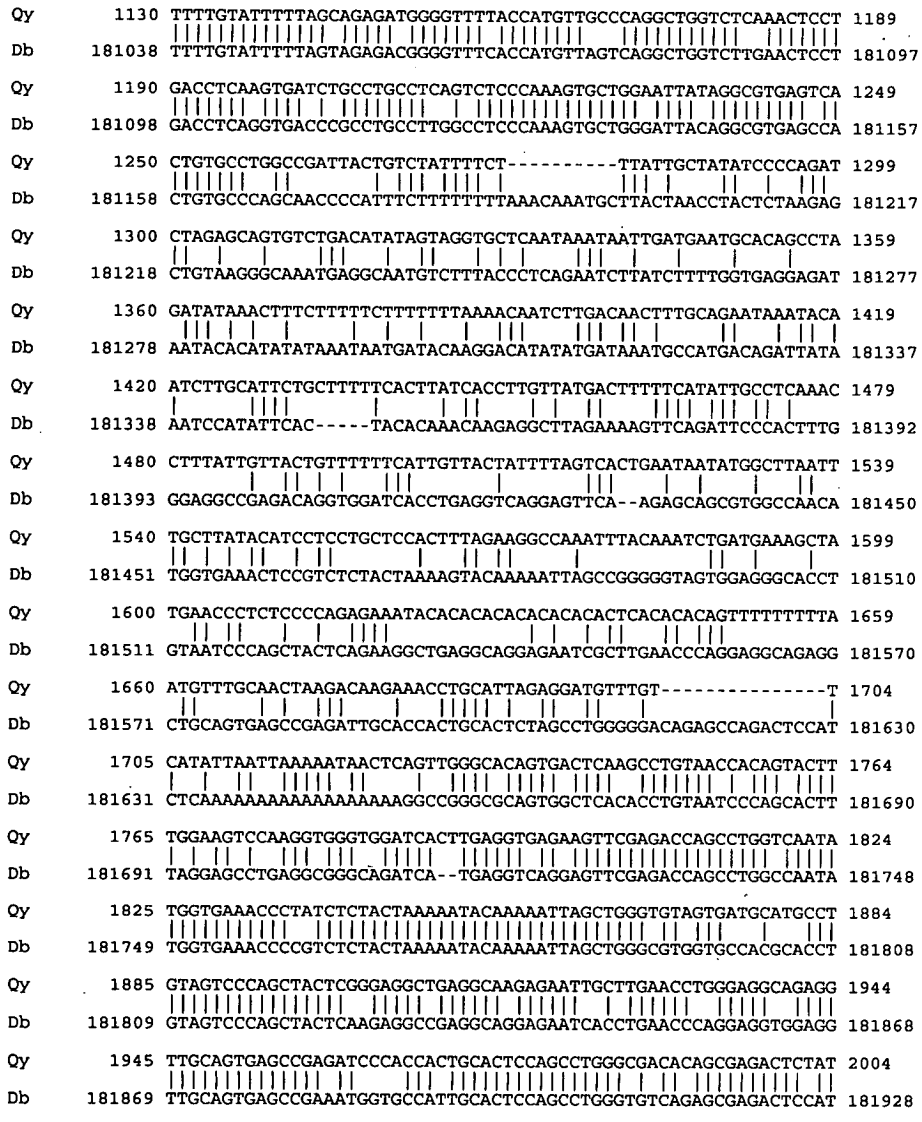
Qy       844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db    180742 GTGATCCTCTCGCCTCATCCTCCCAAAGTGCTGGGATTAGAGGTGTGAGCCACCATGCCC 180801

Qy       904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db    180802 AGCCCATTTTTTTTTAATTTCTTAATTTAATTTCTATTTTTAATTTTTA---TTATTTTA 180858

Qy       964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
Db    180859 TTATTTTTATTTTATTTATTTATATTTTTGAGACAGAGTCTCACTCCGTCGCACAGGCTG 180918

Qy      1024 CCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCT 1069
Db    180919 GAGTGCAATGGTGTGATCTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCAACTCC 180978

Qy      1070 CCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATT 1129
Db    180979 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGGCCTCCACTACACCTGGCTAA-T 181037
```



```
RESULT 14
AC005052/c
LOCUS       AC005052              134210 bp    DNA     linear   PRI 08-NOV-2000
DEFINITION  Homo sapiens BAC clone CTB-38K21 from Xq23, complete sequence.
ACCESSION   AC005052
VERSION     AC005052.2  GI:10122134
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 134210)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 134210)
  AUTHORS   Tin-Wollam,A., Graves,T. and Cofman,M.
  TITLE     The sequence of Homo sapiens BAC clone CTB-38K21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 134210)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 14, 2000 this sequence version replaced gi:3212939.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG038K21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the chromosome X mapping group
            at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/

            SOURCE INFORMATION:
            Clone CTB-38K21 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP3-327A19, 200 base pair
```

```
          overlap.  Actual start of this clone is at base position 1 of
          CTB-38K21; actual end is at base position 9416 of RP3-327A19.
FEATURES             Location/Qualifiers
     source          1. .134210
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq23"
                     /clone="CTB-38K21"
                     /clone_lib="CITB-978SK-B"
     repeat_region   20. .268
                     /rpt_family="Alu"
     repeat_region   269. .561
                     /rpt_family="Alu"
     repeat_region   978. .1285
                     /rpt_family="Alu"
     misc_feature    2964. .4303
                     /note="CpG_island (%GC=68.3, o/e=0.70, #CpGs=110)"
     misc_feature    3597. .3895
                     /note="similar to  EST BE299003 (NID:g9182751)"
     repeat_region   6613. .6919
                     /rpt_family="Alu"
     repeat_region   7154. .7214
                     /rpt_family="L2"
     repeat_region   7640. .7924
                     /rpt_family="Alu"
     repeat_region   8152. .8333
                     /rpt_family="MER1_type"
     repeat_region   8565. .8749
                     /rpt_family="MaLR"
     repeat_region   8750. .9044
                     /rpt_family="Alu"
     repeat_region   9045. .9256
                     /rpt_family="MaLR"
     repeat_region   9271. .9700
                     /rpt_family="L2"
     repeat_region   10238. .10290
                     /rpt_family="L2"
     repeat_region   10291. .10585
                     /rpt_family="Alu"
     repeat_region   10586. .10666
                     /rpt_family="L2"
     repeat_region   10667. .10822
                     /rpt_family="Alu"
     repeat_region   10823. .11015
                     /rpt_family="L2"
     repeat_region   11016. .11199
                     /rpt_family="MER1_type"
     repeat_region   11200. .11309
                     /rpt_family="L2"
     repeat_region   11310. .11544
                     /rpt_family="Alu"
     repeat_region   11545. .11610
                     /rpt_family="L2"
     repeat_region   11860. .12166
                     /rpt_family="Alu"
     misc_feature    12126. .12175
                     /note="similar to  EST AI025436 (NID:g3241049) ow27h07.s1"
     repeat_region   12178. .12228
                     /rpt_family="L1"
     repeat_region   12405. .12520
                     /rpt_family="L1"
     repeat_region   12712. .13008
                     /rpt_family="Alu"
     repeat_region   13009. .13315
                     /rpt_family="Alu"
     repeat_region   13603. .13850
                     /rpt_family="ERVL"
     repeat_region   13851. .14147
                     /rpt_family="Alu"
     repeat_region   14148. .14222
                     /rpt_family="ERVL"
     repeat_region   14223. .14526
                     /rpt_family="Alu"
     repeat_region   14527. .14713
                     /rpt_family="ERVL"
     repeat_region   14815. .15244
                     /rpt_family="L2"
     repeat_region   15245. .15566
                     /rpt_family="Alu"
     repeat_region   15567. .16446
                     /rpt_family="L2"
     repeat_region   16447. .16580
                     /rpt_family="Alu"
     repeat_region   16581. .16866
                     /rpt_family="L2"
     repeat_region   16867. .17178
                     /rpt_family="Alu"
     repeat_region   17179. .17251
                     /rpt_family="L2"
     repeat_region   17255. .17390
                     /rpt_family="MER1_type"
     repeat_region   17872. .18165
                     /rpt_family="Alu"
     repeat_region   18239. .18532
                     /rpt_family="Alu"
     repeat_region   18540. .18841
                     /rpt_family="Alu"
     misc_feature    18631. .18876
                     /note="similar to  EST AW867673 (NID:g8001829)"
     repeat_region   19286. .19570
                     /rpt_family="L2"
     repeat_region   19885. .20185
                     /rpt_family="Alu"
     repeat_region   21149. .21446
                     /rpt_family="Alu"
     repeat_region   21559. .21803
                     /rpt_family="Alu"
     repeat_region   21835. .21951
                     /rpt_family="L1"
     repeat_region   21979. .22286
                     /rpt_family="Alu"
     repeat_region   23260. .23565
                     /rpt_family="Alu"
     repeat_region   23568. .23864
                     /rpt_family="Alu"
     misc_feature    23875. .23924
                     /note="similar to  EST AI735654 (NID:g5057178) at19b08.x1"
     misc_feature    23875. .23924
                     /note="similar to  EST AW619972 (NID:g7326156)"
     misc_feature    23878. .23926

  Query Match            18.0%;  Score 449.2;  DB 8;  Length 134210;
  Best Local Similarity  59.5%;  Pred. No. 3.5e-80;
  Matches  896;  Conservative    1;  Mismatches  554;  Indels   55;  Gaps    6;

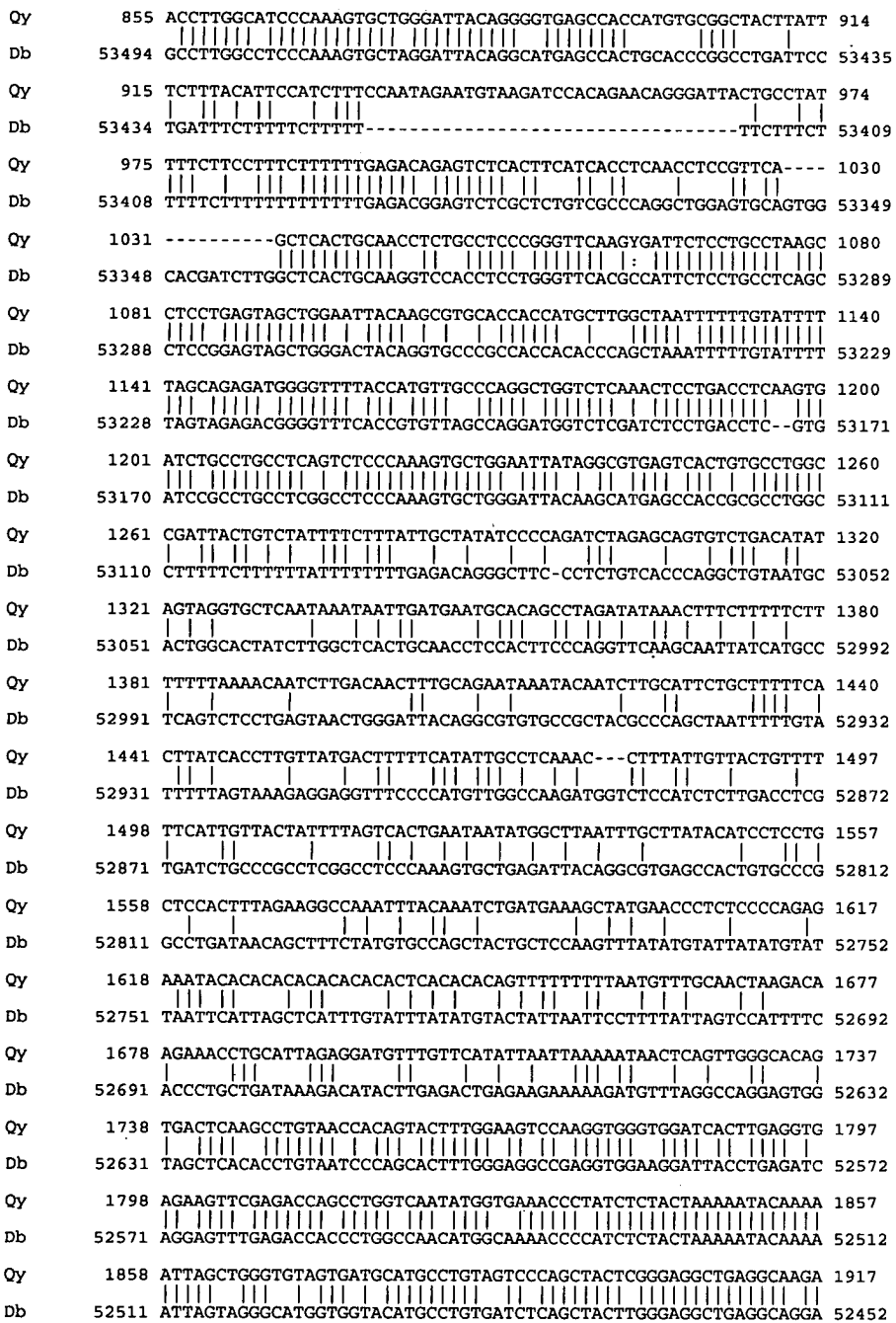
Qy       555 TGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGT 614
             || | |||||||   | |          | |   |||| |  |  ||    ||| ||| |
Db     53793 TGTGGTATTTGCTTTATTTATTTATTTATTTTTTTGGTTAGGGGGTGCTGATTTCTTTTT 53734

Qy       615 TGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCAT 674
             | || ||| |||||||||||| |||||||||||||  |||  |||||   | | |||
Db     53733 TCTTTTTTTGAGACAGGGTCTTTCTCTGTCACCCAGGCTGGAGTGCAGTGACAC-ATCAT 53675

Qy       675 AGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGT 734
               |||||||||||||||||||||| |||||||  |||  |  | ||||||||||| |||
Db     53674 GGCTCACTGCAGCCTCAACCTCCTGGGCTCAAGCAATCCTCAAATCTCAGCCTCCCGAGT 53615

Qy       735 AGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAG 794
             ||||||| |||   || ||  | |||| ||| |||||||||   ||| | ||||
Db     53614 AGCTGGGTCTATAGGCTTGAGCTACCAGGCCCGGCTAATTTTTAAATTGTTTTGTAGAGA 53555

Qy       795 ACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCT 854
               |||||| |||| ||   ||||| |||||||||||||||| ||||||  ||||||||
Db     53554 CAAGGGTCTCACTATGTCACCCAGGGTTGTCTTAAACTCCTGGGTTCAAGGGATCCTCCT 53495
```





RESULT 15
AC005484/c
LOCUS AC005484 131943 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP5-847O8 from 14q24.3, complete sequence.
ACCESSION AC005484
VERSION AC005484.2 GI:5091654
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131943)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 131943)
AUTHORS Cloud,J., Wohldmann,P. and Holmes,A.
TITLE The sequence of Homo sapiens PAC clone RP5-847O8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131943)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 131943)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 131943)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jun 17, 1999 this sequence version replaced gi:3907511.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ0847O8

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

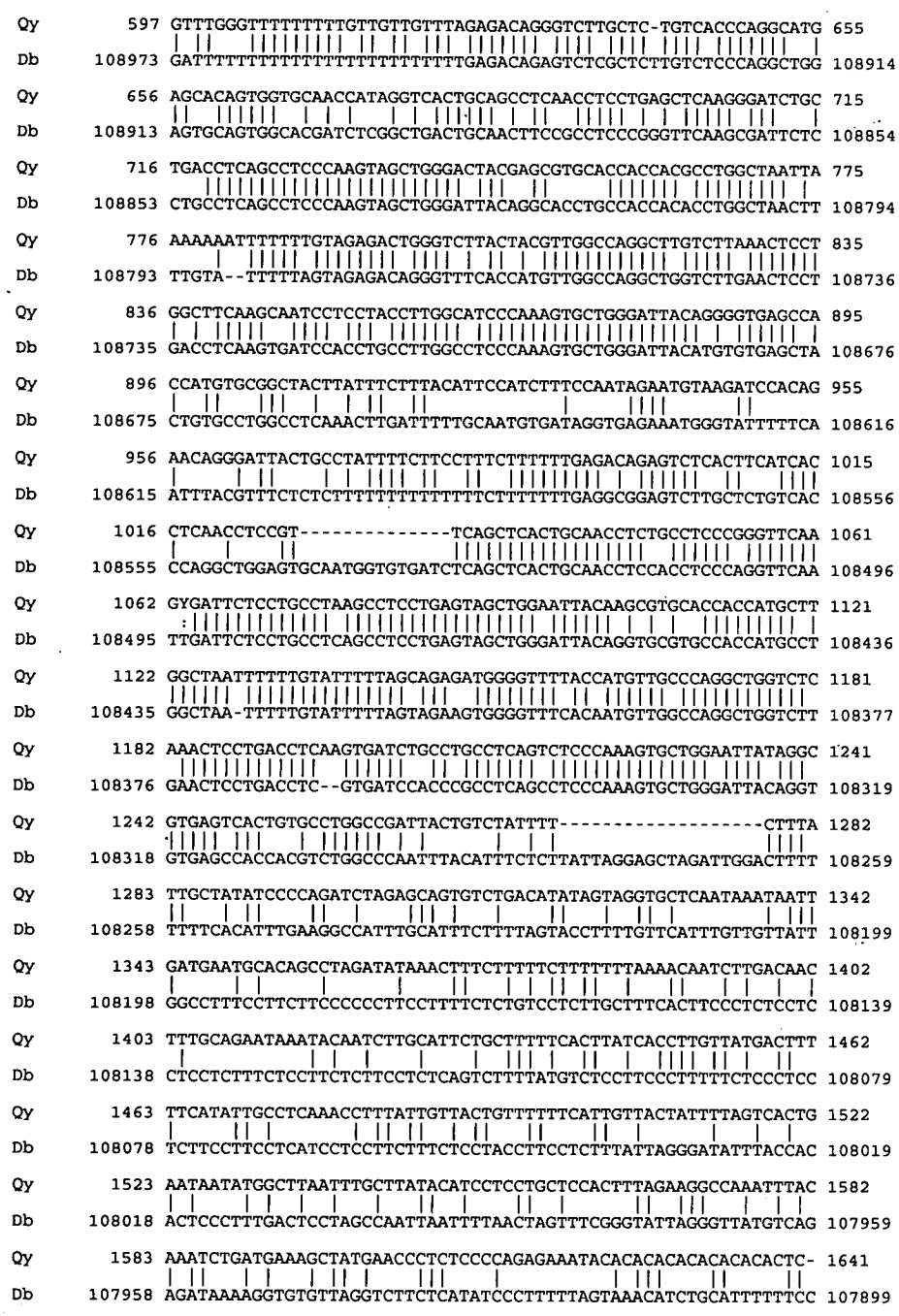
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-892G5; the clone sequenced to the right is RP4-592G7, 200 bp overlap. Actual start of this clone is at base position 1 of RP5-847O8; actual end is at base position 131747 of RP5-847O8.

```
FEATURES             Location/Qualifiers
     source          1. .131943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q24.3"
                     /clone="RP5-847O8"
                     /clone_lib="RPCI-5"
     repeat_region   1058. .1098
                     /rpt_family="AT_rich"
     repeat_region   1762. .2225
                     /rpt_family="MaLR"
     repeat_region   2305. .2587
                     /rpt_family="Alu"
     repeat_region   2937. .3086
                     /rpt_family="L2"
     repeat_region   3293. .3591
                     /rpt_family="Alu"
     repeat_region   3651. .3916
                     /rpt_family="L1"
     repeat_region   3917. .3943
                     /rpt_family="(TA)n"
     repeat_region   4815. .4841
                     /rpt_family="AT_rich"
     repeat_region   5153. .5244
                     /rpt_family="L1"
     repeat_region   5728. .6008
                     /rpt_family="Alu"
     repeat_region   6009. .6034
                     /rpt_family="(CAA)n"
     misc_feature    8080. .8468
                     /note="match to EST R05271 (NID:g755891) ye91b03.s1"
     STS             8080. .8277
     repeat_region   8284. .8589
                     /rpt_family="Alu"
     misc_feature    8649. .9039
                     /note="match to EST R05328 (NID:g755948) ye91b03.r1"
     repeat_region   8857. .9007
                     /rpt_family="MIR"
     repeat_region   9062. .9152
                     /rpt_family="MIR"
     repeat_region   9216. .9288
                     /rpt_family="MIR"
     repeat_region   9476. .9770
                     /rpt_family="Alu"
     STS             9711. .10034
     repeat_region   10004. .10114
                     /rpt_family="Alu"
     repeat_region   10115. .10135
                     /rpt_family="(TAAA)n"
     repeat_region   10136. .10158
                     /rpt_family="Alu"
     repeat_region   10205. .10501
                     /rpt_family="Alu"
     repeat_region   10513. .10819
                     /rpt_family="Alu"
     repeat_region   10890. .11242
                     /rpt_family="MaLR"
     repeat_region   11246. .11519
                     /rpt_family="Alu"
     repeat_region   11567. .11996
                     /rpt_family="L2"
     repeat_region   12006. .12113
                     /rpt_family="L2"
     repeat_region   12123. .12146
                     /rpt_family="(CAAAA)n"
     repeat_region   12147. .12430
                     /rpt_family="Alu"
     repeat_region   12445. .12535
                     /rpt_family="L1"
     repeat_region   12527. .13097
                     /rpt_family="L1"
     repeat_region   13217. .13339
                     /rpt_family="Alu"
     repeat_region   13389. .13994
                     /rpt_family="L2"
     repeat_region   14138. .14158
                     /rpt_family="AT_rich"
     repeat_region   14195. .15071
                     /rpt_family="L1"
     repeat_region   15048. .15521
                     /rpt_family="L1"
     repeat_region   15520. .15992
                     /rpt_family="L1"
     repeat_region   15989. .16221
                     /rpt_family="L1"
     repeat_region   16257. .16556
                     /rpt_family="Alu"
     repeat_region   16569. .16963
                     /rpt_family="L1"
     repeat_region   17016. .17039
                     /rpt_family="AT_rich"
     repeat_region   17073. .17182
                     /rpt_family="MIR"
     repeat_region   17252. .17452
                     /rpt_family="MaLR"
     repeat_region   17473. .17991
                     /rpt_family="L2"
     repeat_region   18106. .18192
                     /rpt_family="MER1_type"
     repeat_region   18193. .18504
                     /rpt_family="Alu"
     repeat_region   18505. .18588
                     /rpt_family="MER1_type"
     repeat_region   18624. .18904
                     /rpt_family="L1"
     repeat_region   19144. .19368
                     /rpt_family="Alu"
     repeat_region   20030. .20159
                     /rpt_family="Alu"
     repeat_region   20167. .20317
                     /rpt_family="Alu"
     repeat_region   20318. .20617
                     /rpt_family="Alu"
     repeat_region   20618. .20753
                     /rpt_family="Alu"
     repeat_region   20802. .20954
                     /rpt_family="L1"
     repeat_region   20955. .21256
                     /rpt_family="Alu"

Query Match             17.9%;  Score 447.8;  DB 8;  Length 131943;
Best Local Similarity   61.3%;  Pred. No. 6.8e-80;
Matches  913;  Conservative    1;  Mismatches  523;  Indels   52;  Gaps   10;
```



Search completed: February 9, 2006, 20:42:48
Job time : 12476.7 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 2691.71 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 499.6
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	499.6	100.0	122302	8	AC003982	AC003982 Homo sapi
c	2	499.6	100.0	220384	14	AC078926	AC078926 Homo sapi
c	3	218.6	43.8	150453	8	AC015920	AC015920 Homo sapi
	4	218.6	43.8	160254	14	AC007763	AC007763 Homo sapi
	5	218.6	43.8	163229	8	AC115085	AC115085 Homo sapi
c	6	218.6	43.8	178097	14	AC016704	AC016704 Homo sapi
c	7	217.2	43.5	79223	8	AL391114	AL391114 Human DNA
c	8	216.4	43.3	169566	8	AC026470	AC026470 Homo sapi
	9	214.8	43.0	662	6	AX386112	AX386112 Sequence
	10	214.8	43.0	123576	8	AC015969	AC015969 Homo sapi
	11	214.8	43.0	190306	8	AC146080	AC146080 Pan trogl
c	12	214.6	43.0	90000	8	AP002789	AP002789 Homo sapi
c	13	214.4	42.9	726	10	BV632439	BV632439 S215P6037
c	14	214.4	42.9	193369	14	AC114889	AC114889 Pan trogl
	15	214.4	42.9	195782	8	AC074389	AC074389 Homo sapi
	16	214.2	42.9	61741	14	AC147308_3	Continuation (4 of
	17	214.2	42.9	110000	14	AC138525_0	AC138525 Homo sapi
	18	214.2	42.9	110000	14	AC147308_2	Continuation (3 of
c	19	214.2	42.9	135758	8	AL136324	AL136324 Human DNA
	20	214.2	42.9	191540	8	AC144780	AC144780 Pan trogl
	21	214.2	42.9	203842	8	AC010789	AC010789 Homo sapi
	22	214	42.8	80068	8	AC008598	AC008598 Homo sapi
c	23	214	42.8	121265	14	AF286885	AF286885 Homo sapi
	24	214	42.8	135213	8	AC111149	AC111149 Homo sapi
	25	214	42.8	149387	8	AC016266	AC016266 Homo sapi
	26	214	42.8	174222	14	AP001197	AP001197 Homo sapi
c	27	214	42.8	175553	14	AP001096	AP001096 Homo sapi
c	28	214	42.8	178042	14	AC018745	AC018745 Homo sapi
	29	214	42.8	182884	8	AC007996	AC007996 Homo sapi
	30	214	42.8	189975	8	AC109512	AC109512 Homo sapi
	31	214	42.8	191318	14	AC020670	AC020670 Homo sapi
	32	213.8	42.8	4278	8	HSM804155	AL832844 Homo sapi
	33	213.8	42.8	136713	8	AC112255	AC112255 Homo sapi
c	34	213.8	42.8	138589	8	AC092966	AC092966 Homo sapi
c	35	213.8	42.8	139953	8	HSBA504H3	AL121585 Human DNA
c	36	213.8	42.8	186649	8	AL929410	AL929410 Human DNA
c	37	213.8	42.8	209157	8	CNS01DW4	AL136332 Human chr
	38	213.4	42.7	587	6	AX389850	AX389850 Sequence
	39	213.4	42.7	59107	14	HS409J21	Z83824 Homo sapien
c	40	213.4	42.7	89086	8	HS1188J21	AL109963 Human DNA
c	41	213.4	42.7	129266	8	AL731567	AL731567 Human DNA
	42	213.4	42.7	153094	14	AC096715	AC096715 Homo sapi
	43	213.4	42.7	159123	8	AC007739	AC007739 Homo sapi
	44	213.4	42.7	168174	6	AR584462	AR584462 Sequence
	45	213.4	42.7	168273	6	AR584403	AR584403 Sequence

ALIGNMENTS

RESULT 1
AC003982/c
LOCUS AC003982 122302 bp DNA linear PRI 18-MAR-1999
DEFINITION Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION AC003982
VERSION AC003982.1 GI:2769695
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 122302)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 122302)
AUTHORS Bradshaw,H., Wu,X. and Ozersky,P.
TITLE The sequence of Homo sapiens PAC clone 166H1
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 122302)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (13-JAN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 122302)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

FEATURES	Location/Qualifiers
source	1. .122302
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="12"
	/map="12q"
	/clone="166H1"
	/clone_lib="RPCI-1"
repeat_region	1. .238
	/rpt_family="Alu"
repeat_region	239. .265
	/rpt_family="AT_rich"
repeat_region	268. .565
	/rpt_family="Alu"
repeat_region	1001. .1309
	/rpt_family="Alu"
repeat_region	1440. .1745
	/rpt_family="Alu"
repeat_region	1909. .1940
	/rpt_family="MER1_type"
repeat_region	1955. .1983
	/rpt_family="AT_rich"
repeat_region	2063. .2362
	/rpt_family="Alu"
repeat_region	2516. .2650
	/rpt_family="Alu"
repeat_region	2674. .2730
	/rpt_family="L1"
repeat_region	2735. .2788
	/rpt_family="(GAAAA)n"
repeat_region	2790. .3077
	/rpt_family="Alu"
repeat_region	3081. .3139
	/rpt_family="L1"
repeat_region	3716. .4020
	/rpt_family="Alu"
repeat_region	4196. .4492
	/rpt_family="Alu"
gene	4560. .14034
	/gene="WUGSC:H_166H1.1"
CDS	join(4560. .5056,13418. .13748,13882. .14034)
	/gene="WUGSC:H_166H1.1"
	/note="unknown function; 60% similar to Z50177 (PID:g927403) (PID:g927402); H_166H1.1"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAB95634.1"
	/db_xref="GI:2769696"
	/translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
repeat_region	5366. .5655
	/rpt_family="Alu"
repeat_region	5672. .5798
	/rpt_family="Alu"
repeat_region	5802. .5834
	/rpt_family="(TA)n"
repeat_region	5834. .5857
	/rpt_family="POLY_A"
repeat_region	5870. .6184
	/rpt_family="Alu"
repeat_region	6224. .6268
	/rpt_family="(CA)n"
repeat_region	6269. .6558
	/rpt_family="Alu"
repeat_region	6729. .7029
	/rpt_family="Alu"
repeat_region	7051. .7172
	/rpt_family="Alu"
repeat_region	7192. .7500
	/rpt_family="Alu"
repeat_region	7518. .7820
	/rpt_family="Alu"
repeat_region	8207. .8500
	/rpt_family="Alu"
repeat_region	8576. .8878
	/rpt_family="Alu"
repeat_region	9023. .9154
	/rpt_family="Alu"
repeat_region	9165. .9470
	/rpt_family="Alu"
repeat_region	9887. .10188
	/rpt_family="Alu"
repeat_region	10221. .10342
	/rpt_family="L1"
repeat_region	10408. .10707
	/rpt_family="Alu"
repeat_region	10713. .11014
	/rpt_family="Alu"
repeat_region	11020. .11110
	/rpt_family="Alu"
repeat_region	11146. .11386
	/rpt_family="Alu"
repeat_region	11208. .11386
	/rpt_family="Alu"
repeat_region	11391. .11515
	/rpt_family="Alu"
repeat_region	11568. .11705
	/rpt_family="Alu"
repeat_region	11737. .12032
	/rpt_family="Alu"
repeat_region	12034. .12066
	/rpt_family="7SLRNA"
repeat_region	12075. .12372
	/rpt_family="Alu"
repeat_region	12448. .12753
	/rpt_family="Alu"
repeat_region	12763. .12874

```
                     /rpt_family="L2"
     repeat_region   12999. .13303
                     /rpt_family="Alu"
     misc_feature    complement(13674. .13752)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     misc_feature    complement(13840. .14195)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     repeat_region   14196. .14229
                     /rpt_family="AT_rich"
     repeat_region   14240. .14339
                     /rpt_family="U6"
     repeat_region   14349. .14646
                     /rpt_family="Alu"
     repeat_region   14652. .14940
                     /rpt_family="Alu"

  Query Match             100.0%;  Score 499.6;  DB 8;  Length 122302;
  Best Local Similarity   99.8%;  Pred. No. 1.9e-142;
  Matches  499;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33287 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 33228

Qy         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33227 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 33168

Qy        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33167 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 33108

Qy        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33107 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 33048

Qy        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33047 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 32988

Qy        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32987 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 32928

Qy        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32927 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 32868

Qy        421 TCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480
              |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db      32867 TCCTTCAGGCCTAAACCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 32808

Qy        481 ACCCACATAAGAGGGACTGA 500
              ||||||||||||||||||||
Db      32807 ACCCACATAAGAGGGACTGA 32788


RESULT 2
AC078926/c
LOCUS       AC078926              220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     2156: contig of 2156 bp in length
          *    2157     2256: gap of unknown length
          *    2257     7112: contig of 4856 bp in length
          *    7113     7212: gap of unknown length
          *    7213    13283: contig of 6071 bp in length
          *   13284    13383: gap of unknown length
          *   13384    20660: contig of 7277 bp in length
          *   20661    20760: gap of unknown length
          *   20761    34473: contig of 13713 bp in length
          *   34474    34573: gap of unknown length
          *   34574    44394: contig of 9821 bp in length
          *   44395    44494: gap of unknown length
          *   44495    57126: contig of 12632 bp in length
          *   57127    57226: gap of unknown length
          *   57227    80266: contig of 23040 bp in length
          *   80267    80366: gap of unknown length
          *   80367   109398: contig of 29032 bp in length
          *  109399   109498: gap of unknown length
          *  109499   149812: contig of 40314 bp in length
          *  149813   149912: gap of unknown length
          *  149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match             100.0%;  Score 499.6;  DB 14;  Length 220384;
  Best Local Similarity   99.8%;  Pred. No. 1.8e-142;
  Matches  499;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78325 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 78266

Qy         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78265 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 78206

Qy        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78205 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 78146

Qy        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78145 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 78086

Qy        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78085 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 78026

Qy        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78025 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAAC 77966

Qy        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77965 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 77906

Qy        421 TCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db      77905 TCCTTCAGGCCTAAATCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 77846

Qy        481 ACCCACATAAGAGGGACTGA 500
              ||||||||||||||||||||
Db      77845 ACCCACATAAGAGGGACTGA 77826


RESULT 3
AC015920/c
LOCUS       AC015920               150453 bp    DNA     linear   PRI 01-OCT-2002
DEFINITION  Homo sapiens chromosome 17, clone CTD-3022L24, complete sequence.
ACCESSION   AC015920
VERSION     AC015920.14  GI:23396301
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 150453)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-3022L24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150453)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 150453)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
```

```
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 150453)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 1, 2002 this sequence version replaced gi:22902021.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L858
                Center clone name: 3022_L_24
            --------------.
FEATURES             Location/Qualifiers
     source          1. .150453
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-3022L24"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   475. .691
                     /rpt_family="L1MDb"
     repeat_region   692. .987
                     /rpt_family="AluSg"
     repeat_region   988. .1402
                     /rpt_family="L1MDb"
     repeat_region   1403. .1705
                     /rpt_family="AluSx"
     repeat_region   1706. .2035
                     /rpt_family="L1MDb"
     repeat_region   2094. .2303
                     /rpt_family="L1MB6"
     repeat_region   2304. .2610
                     /rpt_family="AluSx"
     repeat_region   2611. .2821
                     /rpt_family="L1MB6"
     repeat_region   complement(3303. .3581)
                     /rpt_family="AluSq"
     repeat_region   5069. .5318
                     /rpt_family="MIR"
     repeat_region   7277. .7574
                     /rpt_family="AluSc"
     repeat_region   7959. .7989
                     /rpt_family="AT_rich"
     repeat_region   8313. .8722
                     /rpt_family="LTR24B"
     repeat_region   9000. .9296
                     /rpt_family="MER58B"
     repeat_region   complement(9458. .9612)
                     /rpt_family="L2"
     repeat_region   9749. .9902
                     /rpt_family="L2"
     repeat_region   10228. .10510
                     /rpt_family="AluJb"
     repeat_region   10831. .11020
                     /rpt_family="AluJo"
     repeat_region   11269. .11476
                     /rpt_family="MIR"
     repeat_region   complement(13425. .13948)
                     /rpt_family="MER1A"
     repeat_region   complement(14514. .14823)
                     /rpt_family="AluJb"
     repeat_region   15050. .15404
                     /rpt_family="THE1C"
     repeat_region   16466. .16511
                     /rpt_family="(TTA)n"
     repeat_region   17345. .17857
                     /rpt_family="L3"
     repeat_region   17951. .18350
                     /rpt_family="MLT1D"
     repeat_region   complement(18351. .18651)
                     /rpt_family="AluJo"
     repeat_region   18652. .18760
                     /rpt_family="MLT1D"
     repeat_region   complement(19058. .19171)
                     /rpt_family="MIR"
     repeat_region   complement(19395. .19717)
                     /rpt_family="L2"
     repeat_region   19726. .20025
                     /rpt_family="AluSp"
     repeat_region   20305. .20335
                     /rpt_family="AT_rich"
     repeat_region   21725. .21893
                     /rpt_family="L1MEd"
     repeat_region   21934. .22049
                     /rpt_family="FLAM_C"
     repeat_region   22052. .22081
                     /rpt_family="(CAAAA)n"
     repeat_region   22119. .22178
                     /rpt_family="A-rich"
     repeat_region   complement(22217. .22318)
                     /rpt_family="L2"
     repeat_region   22319. .22596
                     /rpt_family="L1MEd"
     repeat_region   22732. .23010
                     /rpt_family="AluJb"
     repeat_region   23039. .23097
                     /rpt_family="A-rich"
     repeat_region   complement(24817. .24931)
                     /rpt_family="L2"
     repeat_region   24967. .25154
                     /rpt_family="MER5A"
     repeat_region   25739. .25776
                     /rpt_family="AT_rich"
     repeat_region   26394. .26626
                     /rpt_family="AluSc"
     repeat_region   complement(26628. .26928)
                     /rpt_family="AluSq"
     repeat_region   complement(27509. .28031)
                     /rpt_family="MER41A"
     repeat_region   29615. .29747
```

```
                  /rpt_family="AluSq/x"
     repeat_region   29748. .30037
                  /rpt_family="AluSx"
     repeat_region   complement(30051. .30397)
                  /rpt_family="MLT1A2"
     repeat_region   30398. .30446
                  /rpt_family="L1ME4A"
     repeat_region   30753. .31064
                  /rpt_family="AluSx"
     repeat_region   31124. .31231
                  /rpt_family="L1ME4A"

  Query Match             43.8%;  Score 218.6;  DB 8;  Length 150453;
  Best Local Similarity   83.5%;  Pred. No. 5.4e-56;
  Matches  248;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              |||||||||      | ||| |||| |||||||||    ||||| ||||||||||||||||
Db      40951 TTAATAGTAATATCATGGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 40892

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | |||||  |||||||||||| ||||  ||||||||||||||||||||||||||||||||
Db      40891 GGCCGAGAAGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 40832

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              |||| | ||||||| |||||||||||||||||||||||| |||||||  | ||||| || |
Db      40831 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGTAAT 40772

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              ||||||||| | |||||||| ||||| |||||||||||||||||||||  |||||| ||
Db      40771 CCCAGCTACTCTGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCTGAGGTTCCA 40712

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACT 361
              ||||||| ||||  |||||||||||||||||||  || || |  |  |  |  ||
Db      40711 GTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACAAAGCGAGATTCTGTCT 40655


RESULT 4
AC007763
LOCUS       AC007763               160254 bp    DNA     linear   HTG 18-NOV-2001
DEFINITION  Homo sapiens clone RP11-110D17, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC007763
VERSION     AC007763.6  GI:16974272
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 160254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-110D17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 18, 2001 this sequence version replaced gi:14030080.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L854
                Center clone name: 110_D_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 21% of reads
                Sequencing vector: Plasmid; n/a; 79% of reads
                Chemistry: Dye-primer-amersham; 18% of reads
                Chemistry: Dye-terminator Big Dye; 82% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156131 bases at least Q40
                Consensus quality: 157776 bases at least Q30
                Consensus quality: 158528 bases at least Q20
                Insert size: 90000; agarose-fp
                Insert size: 159254; sum-of-contigs
                Quality coverage: 17.8 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20714: contig of 20714 bp in length
            *    20715    20814: gap of 100 bp
            *    20815    21860: contig of 1046 bp in length
            *    21861    21960: gap of 100 bp
            *    21961    23124: contig of 1164 bp in length
            *    23125    23224: gap of 100 bp
            *    23225    25014: contig of 1790 bp in length
            *    25015    25114: gap of 100 bp
            *    25115    26473: contig of 1359 bp in length
            *    26474    26573: gap of 100 bp
            *    26574    31375: contig of 4802 bp in length
            *    31376    31475: gap of 100 bp
            *    31476    40242: contig of 8767 bp in length
            *    40243    40342: gap of 100 bp
            *    40343    55414: contig of 15072 bp in length
            *    55415    55514: gap of 100 bp
            *    55515    76039: contig of 20525 bp in length
            *    76040    76139: gap of 100 bp
            *    76140   110758: contig of 34619 bp in length
            *   110759   110858: gap of 100 bp
            *   110859   160254: contig of 49396 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160254
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-110D17"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .20714
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             20715. .20814
                     /estimated_length=100
     misc_feature    20815. .21860
                     /note="assembly_fragment"
     gap             21861. .21960
                     /estimated_length=100
     misc_feature    21961. .23124
                     /note="assembly_fragment"
     gap             23125. .23224
```

```
                     /estimated_length=100
     misc_feature    23225. .25014
                     /note="assembly_fragment"
     gap             25015. .25114
                     /estimated_length=100
     misc_feature    25115. .26473
                     /note="assembly_fragment"
     gap             26474. .26573
                     /estimated_length=100
     misc_feature    26574. .31375
                     /note="assembly_fragment"
     gap             31376. .31475
                     /estimated_length=100
     misc_feature    31476. .40242
                     /note="assembly_fragment"
     gap             40243. .40342
                     /estimated_length=100
     misc_feature    40343. .55414
                     /note="assembly_fragment"
     gap             55415. .55514
                     /estimated_length=100
     misc_feature    55515. .76039
                     /note="assembly_fragment"
     gap             76040. .76139
                     /estimated_length=100
     misc_feature    76140. .110758
                     /note="assembly_fragment"
     gap             110759. .110858
                     /estimated_length=100
     misc_feature    110859. .160254
                     /note="assembly_fragment"
ORIGIN

  Query Match             43.8%;  Score 218.6;  DB 14;  Length 160254;
  Best Local Similarity   83.5%;  Pred. No. 5.3e-56;
  Matches  248;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              ||||||||||     | ||| |||| |||||||||    ||||| ||||||||||||||||
Db      46713 TTAATAGTAATATCATGGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 46772

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | |||||  |||||||||| ||||  ||||||||||||||||||||||||||||||||
Db      46773 GGCCGAGAAGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 46832

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              |||| | ||||||| |||||||||||||||||||||||| |||||||  | ||||| || |
Db      46833 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGTAAT 46892

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              ||||||||| | |||||||| ||||| |||||||||||||||||||||  |||||| ||
Db      46893 CCCAGCTACTCTGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCTGAGGTTCCA 46952

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACT 361
              ||||||| ||||  |||||||||||||||||||  || || |  |  |  |  ||
Db      46953 GTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACAAAGCGAGATTCTGTCT 47009


RESULT 5
AC115085
LOCUS       AC115085               163229 bp    DNA     linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-280F10, complete sequence.
ACCESSION   AC115085
VERSION     AC115085.4  GI:21314912
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 163229)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-280F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163229)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 163229)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 163229)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (01-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 3, 2002 this sequence version replaced gi:21218421.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26095
                Center clone name: 280_F_10
            --------------.
FEATURES             Location/Qualifiers
     source          1. .163229
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-280F10"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   18. .71
                     /rpt_family="MER58A"
     repeat_region   complement(108. .412)
                     /rpt_family="AluSx"
     repeat_region   413. .579
                     /rpt_family="MER58A"
     repeat_region   676. .811
                     /rpt_family="FLAM_C"
     repeat_region   814. .844
                     /rpt_family="AT_rich"
     repeat_region   complement(2166. .2488)
                     /rpt_family="AluJb"
     repeat_region   3244. .3536
                     /rpt_family="AluSg"
     repeat_region   complement(3628. .3818)
                     /rpt_family="Tigger1"
     repeat_region   complement(3819. .4141)
                     /rpt_family="AluSx"
     repeat_region   complement(4143. .4303)
                     /rpt_family="FRAM"
     repeat_region   complement(4304. .5368)
                     /rpt_family="Tigger1"
     repeat_region   5369. .5400
                     /rpt_family="(TC)n"
     repeat_region   complement(5401. .5418)
                     /rpt_family="Tigger1"
     repeat_region   complement(5419. .5665)
                     /rpt_family="AluJb"
     repeat_region   complement(5666. .6204)
                     /rpt_family="Tigger1"
     repeat_region   complement(6205. .6507)
                     /rpt_family="AluSx"
     repeat_region   complement(6508. .6701)
                     /rpt_family="Tigger1"
     repeat_region   complement(6702. .7009)
                     /rpt_family="AluSq"
     repeat_region   complement(7010. .7497)
                     /rpt_family="Tigger1"
     repeat_region   7756. .8073
                     /rpt_family="AluSg"
     repeat_region   8780. .8968
                     /rpt_family="L1MA6"
     repeat_region   9007. .9303
                     /rpt_family="AluSx"
     repeat_region   10034. .10259
                     /rpt_family="MIR"
     repeat_region   complement(10428. .10732)
                     /rpt_family="AluSx"
     repeat_region   11106. .11392
                     /rpt_family="AluSx"
     repeat_region   11402. .11506
                     /rpt_family="(TA)n"
     repeat_region   11510. .11554
                     /rpt_family="LTR66"
     repeat_region   11573. .11611
                     /rpt_family="(TAGA)n"
     repeat_region   complement(11727. .12008)
                     /rpt_family="AluSq"
     repeat_region   complement(12137. .12267)
                     /rpt_family="FLAM_A"
     repeat_region   12355. .12376
                     /rpt_family="AT_rich"
     repeat_region   complement(12936. .13228)
                     /rpt_family="AluSg"
     repeat_region   13378. .13691
                     /rpt_family="L1PB1"
     repeat_region   15066. .15372
                     /rpt_family="AluSq"
     repeat_region   15539. .15843
                     /rpt_family="AluSx"
     repeat_region   16005. .16067
                     /rpt_family="(TTCA)n"
     repeat_region   16113. .16413
                     /rpt_family="AluSx"
     repeat_region   16692. .17434
                     /rpt_family="L1MB2"
     repeat_region   complement(17491. .17782)
                     /rpt_family="AluSc"
     repeat_region   complement(18043. .18201)
                     /rpt_family="MER20"
     repeat_region   complement(18219. .18321)
                     /rpt_family="MER94"
     repeat_region   18327. .18445
                     /rpt_family="L2"
     repeat_region   complement(18446. .18757)
                     /rpt_family="AluY"
     repeat_region   18758. .19077
                     /rpt_family="L2"
     repeat_region   complement(19078. .19376)
                     /rpt_family="AluSg"

  Query Match           43.8%;  Score 218.6;  DB 8;  Length 163229;
  Best Local Similarity  83.5%;  Pred. No. 5.3e-56;
  Matches  248;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              ||||||||||     | ||| |||| |||||||||   ||||| ||||||||||||||||
Db      50158 TTAATAGTAATATCATGGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 50217

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | |||||  |||||||||||| ||||  ||||||||||||||||||||||||||||||||
Db      50218 GGCCGAGAAGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 50277

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              |||| | ||||||| |||||||||||||||||||||| |||||||  | ||||| || |
Db      50278 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGTAAT 50337

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              ||||||||| | |||||||| ||||| ||||||||||||||||||||  |||||| ||
Db      50338 CCCAGCTACTCTGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCTGAGGTTCCA 50397

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACT 361
              ||||||| ||||  |||||||||||||||||||||  || || |  |  |  |  ||
Db      50398 GTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACAAAGCGAGATTCTGTCT 50454


RESULT 6
AC016704/c
LOCUS       AC016704               178097 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-202G12, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC016704
```

```
VERSION     AC016704.5  GI:8954249
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178097)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178097)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:8569846.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0202G12
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 73%
            Sequencing vector: plasmid; 27%
            Chemistry: Dye-primer ET; 73% of reads
            Chemistry: Dye-terminator Big Dye; 27% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167788 bases at least Q40
            Consensus quality: 171101 bases at least Q30
            Consensus quality: 172980 bases at least Q20
            Insert size: 187000; agarose-fp
            Insert size: 176197; sum-of-contigs
            Quality coverage: 3.54 in Q20 bases; agarose-fp
            Quality coverage: 4.10 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1477: contig of 1477 bp in length
            *     1478     1577: gap of unknown length
            *     1578     3566: contig of 1989 bp in length
            *     3567     3666: gap of unknown length
            *     3667     5440: contig of 1774 bp in length
            *     5441     5540: gap of unknown length
            *     5541     8092: contig of 2552 bp in length
            *     8093     8192: gap of unknown length
            *     8193    10799: contig of 2607 bp in length
            *    10800    10899: gap of unknown length
            *    10900    16410: contig of 5511 bp in length
            *    16411    16510: gap of unknown length
            *    16511    21762: contig of 5252 bp in length
            *    21763    21862: gap of unknown length
            *    21863    27078: contig of 5216 bp in length
            *    27079    27178: gap of unknown length
            *    27179    33120: contig of 5942 bp in length
            *    33121    33220: gap of unknown length
            *    33221    39472: contig of 6252 bp in length
            *    39473    39572: gap of unknown length
            *    39573    48260: contig of 8688 bp in length
            *    48261    48360: gap of unknown length
            *    48361    56756: contig of 8396 bp in length
            *    56757    56856: gap of unknown length
            *    56857    65334: contig of 8478 bp in length
            *    65335    65434: gap of unknown length
            *    65435    75670: contig of 10236 bp in length
            *    75671    75770: gap of unknown length
            *    75771    87026: contig of 11256 bp in length
            *    87027    87126: gap of unknown length
            *    87127    99477: contig of 12351 bp in length
            *    99478    99577: gap of unknown length
            *    99578   111227: contig of 11650 bp in length
            *   111228   111327: gap of unknown length
            *   111328   130565: contig of 19238 bp in length
            *   130566   130665: gap of unknown length
            *   130666   152597: contig of 21932 bp in length
            *   152598   152697: gap of unknown length
            *   152698   178097: contig of 25400 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178097
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-202G12"
     misc_feature    1. .1477
                     /note="assembly_name:Contig9"
     gap             1478. .1577
                     /estimated_length=unknown
     misc_feature    1578. .3566
                     /note="assembly_name:Contig10"
     gap             3567. .3666
                     /estimated_length=unknown
     misc_feature    3667. .5440
                     /note="assembly_name:Contig11"
     gap             5441. .5540
                     /estimated_length=unknown
     misc_feature    5541. .8092
                     /note="assembly_name:Contig12"
     gap             8093. .8192
                     /estimated_length=unknown
     misc_feature    8193. .10799
                     /note="assembly_name:Contig13"
     gap             10800. .10899
                     /estimated_length=unknown
     misc_feature    10900. .16410
                     /note="assembly_name:Contig14"
     gap             16411. .16510
                     /estimated_length=unknown
     misc_feature    16511. .21762
                     /note="assembly_name:Contig15"
     gap             21763. .21862
                     /estimated_length=unknown
     misc_feature    21863. .27078
                     /note="assembly_name:Contig16"
     gap             27079. .27178
                     /estimated_length=unknown
     misc_feature    27179. .33120
                     /note="assembly_name:Contig17"
     gap             33121. .33220
                     /estimated_length=unknown
     misc_feature    33221. .39472
                     /note="assembly_name:Contig18"
     gap             39473. .39572
                     /estimated_length=unknown
     misc_feature    39573. .48260
                     /note="assembly_name:Contig19"
     gap             48261. .48360
                     /estimated_length=unknown
     misc_feature    48361. .56756
                     /note="assembly_name:Contig20"
     gap             56757. .56856
                     /estimated_length=unknown
     misc_feature    56857. .65334
                     /note="assembly_name:Contig21"
     gap             65335. .65434
                     /estimated_length=unknown
     misc_feature    65435. .75670
```

```
                     /note="assembly_name:Contig22"
     gap             75671. .75770
                     /estimated_length=unknown
     misc_feature    75771. .87026
                     /note="assembly_name:Contig23
                     clone_end:T7
                     vector_side:left"
     gap             87027. .87126
                     /estimated_length=unknown
     misc_feature    87127. .99477
                     /note="assembly_name:Contig24"
     gap             99478. .99577
                     /estimated_length=unknown
     misc_feature    99578. .111227
                     /note="assembly_name:Contig25"
     gap             111228. .111327
                     /estimated_length=unknown
     misc_feature    111328. .130565
                     /note="assembly_name:Contig26"
     gap             130566. .130665
                     /estimated_length=unknown
     misc_feature    130666. .152597
                     /note="assembly_name:Contig27"
     gap             152598. .152697
                     /estimated_length=unknown
     misc_feature    152698. .178097
                     /note="assembly_name:Contig28"
ORIGIN

  Query Match             43.8%;  Score 218.6;  DB 14;  Length 178097;
  Best Local Similarity   83.5%;  Pred. No. 5.3e-56;
  Matches  248;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              |||||||||||     | ||| |||| |||||||||   ||||| ||||||||||||||||
Db     175360 TTAATAGTAATATCATGGCCGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 175301

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | |||||  ||||||||||| ||||  ||||||||||||||||||||||||||||||||
Db     175300 GGCCGAGAAGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 175241

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              |||| | ||||||| |||||||||||||||||||||||| |||||||  | ||||| || |
Db     175240 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGTGCATGCCTGTAAT 175181

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              ||||||||| | |||||||| ||||| ||||||||||||||||||||  |||||| ||
Db     175180 CCCAGCTACTCTGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCTGAGGTTCCA 175121

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACT 361
              ||||||| ||||  |||||||||||||||||||  || || |  |  |  |  ||
Db     175120 GTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCAACAAAGCGAGATTCTGTCT 175064


RESULT 7
AL391114/c
LOCUS       AL391114               79223 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-332A4 on chromosome 20 Contains
            part of the RBL1 gene for retinoblastoma-like 1 (p107), the RPS3AP3
            gene for ribosomal protein S3A pseudogene 3, the RPS27AP3 gene for
            ribosomal protein S27a pseudogene 3 and a CpG island, complete
            sequence.
ACCESSION   AL391114
VERSION     AL391114.12  GI:10120128
KEYWORDS    HTG; CpG island; RBL1; retinoblastoma; ribosomal protein; RPS27AP3;
            RPS3AP3.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 79223)
  AUTHORS   Holt,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2000 this sequence version replaced gi:9994293.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP11-332A4 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .79223
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-332A4"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-332A4"
     gene            join(complement(AL136172.16:5346. .5568),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93798. .93962),
                     complement(AL365505.15:86445. .89025))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-001"
     gene            join(complement(AL136172.16:5346. .5562),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93670. .93962))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-002"
     mRNA            join(complement(AL136172.16:5346. .5568),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
```

```
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93798. .93962),
                     complement(AL365505.15:86445. .89025))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-001"
                     /product="retinoblastoma-like 1 (p107)"
                     /note="match: ESTs: AI128307.1 BB624672.1 BB870263.1
                     BE332924.1 BE650226.1 BF398744.1 BG387892.1 BM463052.1
                     BQ004524.1 BU058420.1 BU178775.1 BU556640.1 T85305.1
                     match: cDNAs: Em:L14812.1 Em:M74547.1"
mRNA                 join(complement(AL136172.16:5346. .5562),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93670. .93962))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-001"
                     /product="retinoblastoma-like 1 (p107)"
                     /note="match: ESTs: AA312113.1 AL732304.1 AW209660.1
                     AW354050.1 AW951018.1 BE650226.1 BE854544.1 BG387892.1
                     BM463052.1 BQ899505.1 BU178775.1
                     match: cDNAs: BC032247.1"
CDS                  join(complement(AL136172.16:5346. .5501),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93798. .93962),
                     complement(AL365505.15:88855. .89025))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-001"
                     /standard_name="OTTHUMP00000030892"
                     /note="match: proteins: O55081 P28749 Q08999 Q64700
                     Q64701"
                     /codon_start=1
                     /product="retinoblastoma-like 1 (p107)"
                     /protein_id="CAI95151.1"
                     /db_xref="GI:66347743"
                     /translation="MFEDKPHAEGAAVVAAAGEALQALCQELNLDEGSAAEALDDFTA
                     IRGNYSLEGEVTHWLACSLYVACRKSIIPTVGKGIMEGNCVSLTRILRSAKLSLIQFF
                     SKMKKWMDMSNLPQEFRERIERLERNFEVSTVIFKKYEPIFLDIFQNPYEEPPKLPRS
                     RKQRRIPCSVKDLFNFCWTLFVYTKGNFRMIGDDLVNSYHLLLCCLDLIFANAIMCPN
                     RQDLLNPSFKGLPSDFHTADFTASEEPPCIIAVLCELHDGLLVEAKGIKEHYFKPYIS
                     KLFDRKILKGECLLDLSSFTDNSKAVNKEYEEYVLTVGDFDERIFLGADAEEEIGTPR
                     KFTRDTPLGKLTAQANVEYNLQQHFEKKRSFAPSTPLTGRRYLREKEAVITPVASATQ
                     SVSRLQSIVAGLKNAPSDQLINIFESCVRNPVENIMKILKGIGETFCQHYTQSTDEQP
                     GSHIDFAVNRLKLAEILYYKILETVMVQETRRLHGMDMSVLLEQDIFHRSLMACCLEI
                     VLFAYSSPRTFPWIIEVLNLQPFYFYKVIEVVIRSEEGLSRDMVKHLNSIEEQILESL
                     AWSHDSALWEALQVSANKVPTCEEVIFPNNFETGNGGNVQGHLPLMPMSPLMHPRVKE
                     VRTDSGSLRRDMQPLSPISVHERYSSPTAGSAKRRLFGEDPPKEMLMDKIITEGTKLK
                     IAPSSSITAENVSILPGQTLLTMATAPVTGTTGHKVTIPLHGVANDAGEITLIPLSMN
                     TNQESKVKSPVSLTAHSLIGASPKQTNLTKAQEVHSTGINRPKRTGSLALFYRKVYHL
                     ASVRLRDLCLKLDVSNELRRKIWTCFEFTLVHCPDLMKDRHLDQLLLCAFYIMAKVTK
                     EERTFQEIMKSYRNQPQANSHVYRSVLLKSIPREVVAYNKNINDDFEMIDCDLEDATK
                     TPDCSSGPVKEERGDLIKFYNTIYVGRVKSFALKYDLANQDHMMDAPPLSPFPHIKQQ
                     PGSPRRISQQHSIYISPHKNGSGLTPRSALLYKFNGSPSKSLKDINNMIRQGEQRTKK
                     RVIAIDSDAESPAKRVCQENDDVLLKRLQDVVSERANH"
CDS                  join(complement(AL136172.16:5346. .5501),
                     complement(77685. .77818),complement(56682. .56882),
                     complement(55929. .55993),complement(55689. .55817),
                     complement(55420. .55580),complement(54120. .54169),
                     complement(50780. .50966),complement(49799. .49965),
                     complement(44842. .44954),complement(44249. .44352),
                     complement(35749. .35886),complement(32782. .32946),
                     complement(28849. .28981),complement(23938. .24204),
                     complement(21361. .21572),complement(11346. .11522),
                     complement(9355. .9426),complement(6975. .7065),
                     complement(AL365505.15:97507. .97655),
                     complement(AL365505.15:93789. .93962))
                     /gene="RBL1"
                     /locus_tag="RP11-382A12.1-001"
                     /standard_name="OTTHUMP00000030893"
                     /codon_start=1
                     /product="retinoblastoma-like 1 (p107)"
                     /protein_id="CAI95152.1"
                     /db_xref="GI:66347744"
                     /translation="MFEDKPHAEGAAVVAAAGEALQALCQELNLDEGSAAEALDDFTA
                     IRGNYSLEGEVTHWLACSLYVACRKSIIPTVGKGIMEGNCVSLTRILRSAKLSLIQFF
                     SKMKKWMDMSNLPQEFRERIERLERNFEVSTVIFKKYEPIFLDIFQNPYEEPPKLPRS
                     RKQRRIPCSVKDLFNFCWTLFVYTKGNFRMIGDDLVNSYHLLLCCLDLIFANAIMCPN
                     RQDLLNPSFKGLPSDFHTADFTASEEPPCIIAVLCELHDGLLVEAKGIKEHYFKPYIS
                     KLFDRKILKGECLLDLSSFTDNSKAVNKEYEEYVLTVGDFDERIFLGADAEEEIGTPR
                     KFTRDTPLGKLTAQANVEYNLQQHFEKKRSFAPSTPLTGRRYLREKEAVITPVASATQ
                     SVSRLQSIVAGLKNAPSDQLINIFESCVRNPVENIMKILKGIGETFCQHYTQSTDEQP
                     GSHIDFAVNRLKLAEILYYKILETVMVQETRRLHGMDMSVLLEQDIFHRSLMACCLEI
                     VLFAYSSPRTFPWIIEVLNLQPFYFYKVIEVVIRSEEGLSRDMVKHLNSIEEQILESL
                     AWSHDSALWEALQVSANKVPTCEEVIFPNNFETGNGGNVQGHLPLMPMSPLMHPRVKE
                     VRTDSGSLRRDMQPLSPISVHERYSSPTAGSAKRRLFGEDPPKEMLMDKIITEGTKLK
                     IAPSSSITAENVSILPGQTLLTMATAPVTGTTGHKVTIPLHGVANDAGEITLIPLSMN
                     TNQESKVKSPVSLTAHSLIGASPKQTNLTKAQEVHSTGINRPKRTGSLALFYRKVYHL
```

Query Match 43.5%; Score 217.2; DB 8; Length 79223;
Best Local Similarity 85.0%; Pred. No. 1.5e-55;
Matches 243; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Qy        60 CTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTT 119
             |||    ||    |||  ||||| | |||||  ||||||||    |||||  ||||||||||
Db     64472 CTCTAGAAAAGTGAACAAATTAGGCCGGTGCAGTGGCTCACACCTGTAATCCCAGCACTT 64413

Qy       120 TGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACA 179
             ||||||  |||||  ||||||||||||  ||||  ||||||||||||||||  ||||||||||
Db     64412 TGGGAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCACCCTGGCCAACA 64353

Qy       180 TGGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCT 239
             |||||||||  | ||||||||  ||||||||||||||||  ||||||||||||||  |||||
Db     64352 TGGCAAAACCCCATCTCTACCAAAAATACAAAAATTAGTCAGGTGTGGTGGCATGTGCCT 64293

Qy       240 ATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGG 299
             |||  ||||||||||  ||||||||||  ||  ||  ||||||||||||||||  ||||||||||
Db     64292 ATAATCCCAGCTACTCAGGAGGCTGAGGGAGGAGAATCACTTGAACCCAGGAGGTGGAGG 64233

Qy       300 TTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAA 345
             |||||||||||||||||  ||  ||||||||||  |||||||  | | |
Db     64232 TTGCAGTGAGCCAAGATCGCGCCACTGCACTCTAGCCTGGGAGACA 64187
```

```
RESULT 8
AC026470/c
LOCUS       AC026470              169566 bp    DNA     linear   PRI 25-NOV-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-44I10, complete sequence.
ACCESSION   AC026470
VERSION     AC026470.6  GI:38502331
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 169566)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169566)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 169566)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 169566)
  AUTHORS   Stanford Human Genome Center and Los Alamos National Laboratory.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 25, 2003 this sequence version replaced gi:19072800.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .169566
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-44I10"
ORIGIN

  Query Match           43.3%;  Score 216.4;  DB 8;  Length 169566;
  Best Local Similarity   77.5%;  Pred. No. 2.5e-55;
  Matches  262;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy         30 TAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTG 89
              |||| |  |  ||   |   |   | | | ||  |  ||   | |   ||| ||||||| |
Db      84067 TAGCCTGGGCAACAGAGTGAGACTTCATCTCTATGAAAAACTTTAAAAATTGGCCAGGCG 84008

Qy         90 CGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAG 149
              |||||||||||| ||||| ||||||| ||||||||| |||||  ||| ||||||||||||
Db      84007 TGGTGGCTCATGCCTGTAATCCCAGCGCTTTGGGAGGCCGAGGCGGGCGGATCACTTGAG 83948

Qy        150 AGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACA 209
              |||||||| |||| |||||||| ||||||||||||||| | |||||||| ||||||||||
Db      83947 GTCAGGAGTTAGAGATCAGCCTGGTCAACATGGCAAAACCCCATCTCTATTAAAAATACA 83888

Qy        210 AAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCA 269
              |||||||||  ||  |||| || |  |||| || ||||||||||   |||||||| ||||
Db      83887 AAAATTAGCTGGGCATGGTTGCCCACGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCA 83828

Qy        270 GAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCAC 329
              | |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      83827 GGAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATCGCACCACTGCAC 83768

Qy        330 TCCAGCCTGGAAAAAAAGGGTAATTAATAACTTTACTT 367
              ||||||||||  || |  |  |||   |  ||| | ||
Db      83767 TCCAGCCTGGGCAACAGAGTGAATCTGTGTCTTAAATT 83730


RESULT 9
AX386112
LOCUS       AX386112                 662 bp    DNA     linear   PAT 19-MAR-2002
DEFINITION  Sequence 1040 from Patent WO0214500.
ACCESSION   AX386112
VERSION     AX386112.1  GI:19579242
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Escobedo,J., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C.,
            Randazzo,F., Lamson,G., Scott,E.M., Zhang,G., Kassam,A., Pot,D. and
            Labat,I.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0214500-A 1040 21-FEB-2002;
            CHIRON CORPORATION (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           43.0%;  Score 214.8;  DB 6;  Length 662;
  Best Local Similarity   83.8%;  Pred. No. 1e-54;
  Matches  243;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         63 AGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGG 122
              ||| || |||     ||| ||||||  |||||||||   ||||| ||||||||||||||
Db        184 AGTAAAGAGTTTTGGATTGGCCAGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGG 243

Qy        123 GAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGG 182
              ||| |||||   ||||||||||||||  ||||||| |||||||| ||||||||||||||
Db        244 GAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGG 303

Qy        183 CAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
              ||||  || ||||||||||||||||||||||||||||||||||||||||  |||| ||
Db        304 TGAAACCCTGTCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGCGCCTGTA 363

Qy        243 GTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTG 302
              | ||||||||| | || ||||| ||||| |||||| |||||||||||||||||||| ||
Db        364 GCCCCAGCTACTCTGGCGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGCTG 423

Qy        303 CAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||  |||| ||||||||||| ||||||||||  || ||  || |
Db        424 CAGTGAGCTGAGATCGCACCACTGCATTCCAGCCTGGGCAACAAAAGTGA 473


RESULT 10
AC015969
LOCUS       AC015969               123576 bp    DNA     linear   PRI 16-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-218N6 from 2, complete sequence.
ACCESSION   AC015969
VERSION     AC015969.9  GI:15145265
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 123576)
  AUTHORS   Cotton,M., Doebber,A., Mulvaney,E. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-218N6
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 123576)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 123576)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Genome Sequencing Center, Washington
```

```
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 123576)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 123576)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 123576)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14488387.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0218N06
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-474G23, 2000 bp overlap.
            Actual start of this clone is at base position 144971 of
            RP11-474G23; actual end is at base position 123576 of RP11-218N6.

            RP11-218N6 contains a single plasmid region from 42433 to 42444.
FEATURES             Location/Qualifiers
     source          1. .123576
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-218N6"
                     /clone_lib="RPCI-11"
     gene            complement(16243. .42944)
                     /gene="DKFZP566K1924"
     mRNA            complement(join(16243. .17259,40390. .40540,42455. .42944))
                     /gene="DKFZP566K1924"
     CDS             complement(join(17095. .17259,40390. .40540,42455. .42633))
                     /gene="DKFZP566K1924"
                     /note="Homo sapiens DKFZP566K1924 protein (DKFZP566K1924),
                     mRNA.; H_NH0218N06.1
                     This gene was based on gi(24308070)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAX93120.1"
                     /db_xref="GI:62702194"
                     /translation="MGDLPGLVRLSIALRIQPNDGPVFYKVDGQRFGQNRTIKLLTGS
                     SYKVEVKIKPSTLQVENISIGGVLVPLELKSKEPDGDRVVYTGTYDTEGVTPTKSGER
                     QPIQITMPFTDIGTFETVWQVKFYNYHKRDHCQWGSPFSVIEYECKPNETRSLMWVNK
                     ESFL"
     misc_feature    26142. .26379
                     /note="CpG_island (%GC=61.3, o/e=0.90, #CpGs=19)"
     misc_feature    42441. .43212
                     /note="CpG_island (%GC=68.8, o/e=0.74, #CpGs=75)"
     misc_feature    77919. .78140
                     /note="CpG_island (%GC=60.4, o/e=0.80, #CpGs=16)"
     gene            88517. .120686
                     /gene="PLEK"
     mRNA            join(88517. .88626,103561. .103716,103956. .104137,
                     105775. .105866,109735. .109919,111620. .111724,
                     116395. .116478,117340. .117409,118913. .120686)
                     /gene="PLEK"
     CDS             join(88585. .88626,103561. .103716,103956. .104137,
                     105775. .105866,109735. .109919,111620. .111724,
                     116395. .116478,117340. .117409,118913. .119049)
                     /gene="PLEK"
                     /note="Homo sapiens pleckstrin (PLEK), mRNA.;
                     H_NH0218N06.2
                     This gene was based on gi(4505878)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAX93121.1"
                     /db_xref="GI:62702195"
                     /translation="MEPKRIREGYLVKKGSVFNTWKPMWVVLLEDGIEFYKKKSDNSP
                     KGMIPLKGSTLTSPCQDFGKRMFVFKITTTKQQDHFFQAAFLEERDAWVRDIKKAIKC
                     IEGGQKFARKSTRRSIRLPETIDLGALYLSMKDTEKGIKELNLEKDKKIFNHCFTGNC
                     VIDWLVSNQSVRNRQEGLMIASSLLNEGYLQPAGDMSKSAVDGTAENPFLDNPDAFYY
                     FPDSGFFCEENSSDDDVILKEEFRGVIIKQGCLLKQGHRRKNWKVRKFILREDPAYLH
                     YYDPAGAEDPLGAIHLRGCVVTSVESNSNGRKSEEENLFEIITADEVHYFLQAATPKE
                     RTEWIRAIQMASRTGK"
ORIGIN

  Query Match          43.0%;  Score 214.8;  DB 8;  Length 123576;
  Best Local Similarity  83.8%;  Pred. No. 8e-55;
  Matches  243;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         63 AGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGG 122
              ||| || |||      ||| ||||||  ||||||||||    ||||| ||||||||||||||
Db      33261 AGTAAAGAGTTTTGGATTGGCCAGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGG 33320

Qy        123 GAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGG 182
              ||| |||||   |||||||||||||||  |||||||| ||||||||| ||||||||||||
Db      33321 GAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGG 33380

Qy        183 CAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
              |||| || ||||||||||||||||||||||||||||||||||||||||  |||| ||
Db      33381 TGAAACCCTGTCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGCGCCTGTA 33440

Qy        243 GTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTG 302
              | ||||||||| | || ||||| ||||| |||||| ||||||||||||||||||||| ||
Db      33441 GCCCCAGCTACTCTGGCGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGCTG 33500

Qy        303 CAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||  |||| |||||||||| ||||||||||  || ||  || |
Db      33501 CAGTGAGCTGAGATCGCACCACTGCATTCCAGCCTGGGCAACAAAAGTGA 33550


RESULT 11
AC146080
LOCUS       AC146080              190306 bp    DNA     linear   PRI 19-MAY-2004
```

```
DEFINITION  Pan troglodytes BAC clone RP43-51K11 from 7, complete sequence.
ACCESSION   AC146080
VERSION     AC146080.4  GI:47498253
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 190306)
  AUTHORS   Tomlinson,C. and Bielicki,L.
  TITLE     The sequence of Pan troglodytes BAC clone RP43-51K11
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 190306)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 190306)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2004) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 190306)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Washington University School of Medicine,
            Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
            63108, USA
COMMENT     On May 19, 2004 this sequence version replaced gi:46559537.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: C_PT051K11
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
            was isolated from white blood cells obtained from a male chimpanzee
            (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The
            clone and detailed information can be obtained from ResGen
            (http://www.resgen.com) or Pieter de Jong and co-workers at
            http://www.bacpac.chori.org.

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .190306
                        /organism="Pan troglodytes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9598"
                        /chromosome="7"
                        /map="7"
                        /clone="RP43-51K11"
                        /clone_lib="RPCI-43"
ORIGIN

  Query Match             43.0%;  Score 214.8;  DB 8;  Length 190306;
  Best Local Similarity   82.6%;  Pred. No. 7.8e-55;
  Matches  246;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         67 AATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGA 126
              ||||   |   |   |||||||| |||||||||| ||||| |||||||||||||| ||
Db     133566 AATAAAGAAAAAGAGGCCAGGTGAGGTGGCTCATGCCTGTAATCCCAGCACTTTGGAAGG 133625

Qy        127 CCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAA 186
              | ||| |||||||||||||||||  ||||||||  |||||||||||||||| ||||  ||
Db     133626 CTGAGGTGGGTGGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAAGATGGTGAA 133685

Qy        187 ACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCC 246
              || | ||||||  |||||||||||||||||||||||||||||| ||||||| ||||| || |||
Db     133686 ACCCCATCTCTGCTAAAAATACAAAAATTAGCCAGGTGTTGTGGCACATGCCTGTAATCC 133745

Qy        247 CAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGT 306
              ||||||| |||||||||| ||||| |||||||||||||||||||| | ||||||||||||
Db     133746 CAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAAGCGGAGGTTGCAGT 133805

Qy        307 GAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTTA 364
              ||||| |||||||||||||||||||||||||||||    | |  |  |     | |||| |
Db     133806 GAGCCGAGATTGCACCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTGTGACTTAA 133863


RESULT 12
AP002789/c
LOCUS       AP002789               90000 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-716L7,
            complete sequence.
ACCESSION   AP002789
VERSION     AP002789.3  GI:15320481
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 90000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Aug 27, 2001 this sequence version replaced gi:11071948.
FEATURES             Location/Qualifiers
     source          1. .90000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-716L7"
ORIGIN

  Query Match             43.0%;  Score 214.6;  DB 8;  Length 90000;
  Best Local Similarity   79.8%;  Pred. No. 9.3e-55;
  Matches  253;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy         69 TAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACC 128
              ||  | ||  |   ||||| |||||||||| | ||||| |||||||||||||||| ||
Db      70289 TAAGAGCTTCTGCACCAGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCC 70230
```

```
Qy        129 GAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAAC 188
              ||| |||||||||||| ||||  ||||||||||||||||||||||||||||||  ||||
Db      70229 GAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAAC 70170

Qy        189 ACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCA 248
              || |||||| |||||||||||||||||||||| ||| |||||| ||||| ||||||||
Db      70169 CCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCGTGTTGGCACATGCCTGTAGTCCCA 70110

Qy        249 GCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGA 308
              ||||| |||||||||| ||||  |||||  |||||||| ||||||||||||||||||| ||
Db      70109 GCTACTCAGGAGGCTGAGGCAAGAGAATAGCTTGAACCCGGGAGGTGGAGGTTGCAGCGA 70050

Qy        309 GCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTTACTTG 368
              |||||||||  ||||||  ||||||||||||  || | ||  |      | ||  |
Db      70049 GCCAAGATTATACCACTAAACTCCAGCCTGGGCAACATGGTGAGACTCCATCTCAAAAAA 69990

Qy        369 CAACCATAGCTGCTTCT 385
              ||  | |   ||||||
Db      69989 AAAAAAAAAGAGCTTCT 69973
```

RESULT 13
BV632439/c
LOCUS BV632439 726 bp DNA linear STS 15-APR-2005
DEFINITION S215P60370FH7.T0 Clara Pan troglodytes troglodytes STS genomic, sequence tagged site.
ACCESSION BV632439
VERSION BV632439.1 GI:62622617
KEYWORDS STS.
SOURCE Pan troglodytes troglodytes
ORGANISM Pan troglodytes troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 726)
AUTHORS Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and Jaffe,D.B.
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome
JOURNAL Unpublished (2005)
COMMENT
Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 726
Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI
Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan
troglodytes verus), 3 other Pan troglodytes verus chimps (Donald,Karlien,Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps of unknown origin
(Gon,Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base calls must have Phred
score >= 20, at least 30% of its base calls must satisfy SNQS(30,25)(single strand NQS, the
base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred
score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely
placed in the genome and read pairs whose two ends were not consistently placed were
discarded. After above filtering, NQS(30,25) standard was applied to all pairs of
overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less
than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two
copies of a single read, comparisons between two reads that share 95% of their genome
alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus
of human genome) were discarded.
FEATURES Location/Qualifiers
source 1. .726
/organism="Pan troglodytes troglodytes"
/mol_type="genomic DNA"
/sub_species="troglodytes"
/db_xref="taxon:37011"
/clone_lib="Clara"
STS <1. .>726
ORIGIN

Query Match 42.9%; Score 214.4; DB 10; Length 726;
Best Local Similarity 80.4%; Pred. No. 1.4e-54;
Matches 251; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy         46 GGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTG 105
              | ||  | |||| || |    || || | | ||   |||||| | |||||||||   |||
Db        312 GAATTAAATTAGTGCCCTTAAAAAAGGACCCCAGAGGCCAGGAGTGGTGGCTCACACCTG 253

Qy        106 TATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGAC 165
              || |||||||||||||||||  ||||| ||||  ||||||||||  |||||||||||||
Db        252 TAATCCCAGCACTTTGGGAAGCCGAGGTGGGCAGATCACTTGAGGTCAGGAGTTTGAGAC 193

Qy        166 CAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGT 225
              |||||||||||||||||||||| |||| |||| ||||||||||||||||||||||| |||
Db        192 TAGCCTGGCCAACATGGCAAAACCCTATTTCTACTAAAAATACAAAAATTAGCCAGATGT 133

Qy        226 GGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAAC 285
              |||||||| ||||| || | |||||||| |||||||||| ||||  |||||  |||||||
Db        132 GGTGGCACATGCCTTTAATTCCAGCTACTCAGGAGGCTGAGGCATGAGAATTTCTTGAAC 73

Qy        286 CTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAA 345
              |  |||| ||||||||||||  |||||||| |||||||||||||||||||||||    | |
Db         72 CCAGGAGATGGAGGTTGCAGGAAGCCAAGATGGCACCACTGCACTCCAGCCTGGGTGATA 13

Qy        346 AGGGTAATTAAT 357
              | |  |   |||
Db         12 AAGCAAGAGAAT 1
```

RESULT 14
AC114889/c
LOCUS AC114889 193369 bp DNA linear HTG 18-JAN-2003
DEFINITION Pan troglodytes clone RP43-64L14, WORKING DRAFT SEQUENCE, 5 ordered pieces.
ACCESSION AC114889
VERSION AC114889.3 GI:27802022
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 193369)
AUTHORS Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,

```
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193369)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 193369)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jan 18, 2003 this sequence version replaced gi:21622727.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: cld
                 Center clone name: 064L14

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 192665 bases at least Q40
                 Consensus quality: 192877 bases at least Q30
                 Consensus quality: 192937 bases at least Q20
                 Insert size: 120000; agarose-fp
                 Insert size: 192969; sum-of-contigs
                 Quality coverage: 15.56x in Q20 bases; agarose-fp
                 Quality coverage: 9.68x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     2008: contig of 2008 bp in length
            *     2009     2108: gap of unknown length
            *     2109    25979: contig of 23871 bp in length
            *    25980    26079: gap of unknown length
            *    26080   158388: contig of 132309 bp in length
            *   158389   158488: gap of unknown length
            *   158489   188878: contig of 30390 bp in length
            *   188879   188978: gap of unknown length
            *   188979   193369: contig of 4391 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193369
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-64L14"
                     /clone_lib="RP43"
     misc_feature    1. .63570
                     /note="clone overlaps with GenBank Accession Number
                     AC116439 clone RP43-88G1 (center project name clc)"
     misc_feature    1. .2008
                     /note="assembly_fragment
                     missing T7 clone end on 5' end of insert"
     gap             2009. .2108
                     /estimated_length=unknown
     misc_feature    2109. .25979
                     /note="assembly_fragment"
     gap             25980. .26079
                     /estimated_length=unknown
     misc_feature    26080. .158388
                     /note="assembly_fragment"
     gap             158389. .158488
                     /estimated_length=unknown
     misc_feature    158489. .188878
                     /note="assembly_fragment"
     misc_feature    161692. .193369
                     /note="clone overlaps with GenBank Accession Number
                     AC114888 clone RP43-159M22 (center project name cle)"
     gap             188879. .188978
                     /estimated_length=unknown
     misc_feature    188979. .193369
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             42.9%;  Score 214.4;  DB 14;  Length 193369;
  Best Local Similarity   84.0%;  Pred. No. 1e-54;
  Matches  242;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
               |||| | ||    |  |||||||||| |||||||||| ||||| ||||||||||||||||||
Db      121917 TTAAAAATATTCAAACAGCCAGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA 121858

Qy         125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
               |||| ||  |||  |||||| |||   ||||||||||||||||||||||||||||||||
Db      121857 GACCAAGGCGGGCAGATCACCTGAAGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTG 121798

Qy         185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
               |||| || |||||| |||||||||||||||||  |||||||||||| ||||| || |
Db      121797 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCACATGCCTGTAAT 121738

Qy         245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
               |||||||||  ||||||||| ||||| |||||| ||||||||||||||||||||||||||
Db      121737 CCCAGCTACTTAGGAGGCTGAGGCAGGAGAATCTCTTGAACCTGGGAGGTGGAGGTTGCA 121678

Qy         305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
               ||||||| ||||||  ||| || |||||||| |||  || ||| | ||
Db      121677 GTGAGCCGAGATTGTGCCATTGTACTCCAGCTTGGCCAACAAGAGCAA 121630


RESULT 15
AC074389
LOCUS       AC074389              195782 bp    DNA     linear   PRI 08-OCT-2003
DEFINITION  Homo sapiens BAC clone RP11-510K8 from 7, complete sequence.
ACCESSION   AC074389
VERSION     AC074389.8  GI:18042461
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 195782)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 195782)
  AUTHORS   Cordes,M., Doebber,A., Hawkins,M. and Kozlowicz,A.
```

```
  TITLE     The sequence of Homo sapiens BAC clone RP11-510K8
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 195782)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 195782)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 195782)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 195782)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 195782)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 3, 2002 this sequence version replaced gi:14209801.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0510K08
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:

            Data from AC093734 was used to finish this clone, AC074389.

            The clone sequenced to the left is RP11-16P10, 2000 bp overlap.
            Actual start of this clone is at base position 112086 of RP11-16P10
            actual end is at base position 195782 of RP11-510K8.
FEATURES             Location/Qualifiers
     source          1. .195782
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-510K8"
                     /clone_lib="RPCI-11"
     repeat_region   3. .269
                     /rpt_family="L1"
     repeat_region   270. .372
                     /rpt_family="MIR"
     repeat_region   558. .736
                     /rpt_family="Alu"
     repeat_region   807. .844
                     /rpt_family="(TTCA)n"
     repeat_region   946. .1126
                     /rpt_family="C-rich"
     repeat_region   1099. .1278
                     /rpt_family="CT-rich"
     repeat_region   1417. .1471
                     /rpt_family="CT-rich"
     repeat_region   1485. .1595
                     /rpt_family="L1"
     repeat_region   1621. .1647
                     /rpt_family="(CA)n"
     repeat_region   2190. .2260
                     /rpt_family="(GGCTG)n"
     repeat_region   2261. .2383
                     /rpt_family="MIR"
     repeat_region   2804. .3125
                     /rpt_family="L2"
     repeat_region   3175. .3277
                     /rpt_family="MaLR"
     repeat_region   3599. .3761
                     /rpt_family="L2"
     repeat_region   3762. .4059
                     /rpt_family="Alu"
     repeat_region   4060. .4157
                     /rpt_family="L2"
     repeat_region   4173. .4432
                     /rpt_family="L2"
     repeat_region   5572. .5740
                     /rpt_family="MIR"
     repeat_region   5756. .6058
                     /rpt_family="Alu"
     repeat_region   6506. .6651
                     /rpt_family="L2"
     repeat_region   6652. .6934
                     /rpt_family="Alu"
     repeat_region   6935. .7090
                     /rpt_family="L2"
     repeat_region   7091. .7404
                     /rpt_family="Alu"
     repeat_region   7503. .7785
                     /rpt_family="Alu"
     repeat_region   7787. .7815
                     /rpt_family="ERV1"
     repeat_region   7816. .8151
                     /rpt_family="Alu"
     repeat_region   8152. .8534
                     /rpt_family="ERV1"
     repeat_region   8535. .8804
                     /rpt_family="Alu"
     repeat_region   10374. .10668
                     /rpt_family="Alu"
     repeat_region   10854. .11031
```

```
                     /rpt_family="MER1_type"
     repeat_region   11051. .11212
                     /rpt_family="ERV1"
     repeat_region   11235. .11328
                     /rpt_family="MIR"
     repeat_region   11692. .11800
                     /rpt_family="MaLR"
     repeat_region   11801. .12113
                     /rpt_family="Alu"
     repeat_region   12114. .12415
                     /rpt_family="MaLR"
     repeat_region   12559. .13339
                     /rpt_family="(TG)n"
     repeat_region   13357. .13530
                     /rpt_family="(TG)n"
     repeat_region   13564. .13738
                     /rpt_family="(TG)n"
     misc_feature    13740. .14544
                     /note="CpG_island (%GC=69.6, o/e=0.82, #CpGs=65)"
     repeat_region   14474. .14593
                     /rpt_family="MaLR"
     repeat_region   14597. .14909
                     /rpt_family="Alu"
     repeat_region   14920. .15038
                     /rpt_family="L1"
     misc_feature    15062. .15480
                     /note="CpG_island (%GC=65.9, o/e=0.79, #CpGs=33)"
     repeat_region   15536. .15708
                     /rpt_family="L1"
     repeat_region   16108. .16168
                     /rpt_family="Alu"
     repeat_region   16169. .16379
                     /rpt_family="Alu"
     repeat_region   16380. .16429
                     /rpt_family="Alu"
     repeat_region   16430. .16732
                     /rpt_family="Alu"

  Query Match             42.9%;  Score 214.4;  DB 8;  Length 195782;
  Best Local Similarity   84.0%;  Pred. No. 1e-54;
  Matches  242;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              || ||| | | ||    ||| ||||| ||||||||   ||||| ||||||||||||||||
Db      55184 TTTATATTGAGTCCCCGGCCGGGTGCAGTGGCTCACTCCTGTAGTCCCAGCACTTTGGGA 55243

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | | |||  |||||||||||| ||||| |||||||| ||||||||||||||||||||||
Db      55244 GGCTGAGGCGGGTGGATCACCTGAGATCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 55303

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              ||||  |  |||||| |||||||||||||||||||||||||||||||||| |||||||| |
Db      55304 AAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGTGCCTATAAT 55363

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              ||  |||||| | |||||| | ||||| |||||||||||||| |  ||||||||||||||||
Db      55364 CCTAGCTACTCGGGAGGCCGAGGCAGGAGAATCACTTGAATCCAGGAGGTGGAGGTTGCA 55423

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||| |||| |||||||||||||||||||||    | ||| || |
Db      55424 GTGAGCCGAGATCGCACCACTGCACTCCAGCCTGGGGGACAAGAGTGA 55471


Search completed: February  9, 2006, 18:20:22
Job time : 2696.71 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 344.27 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 499.6
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	499.6	100.0	12174	12	ADI35082	Adi35082 Human PLA
	2	499.6	100.0	12174	12	ADJ09983	Adj09983 Human pho
	3	214.8	43.0	662	6	ABN61073	Abn61073 Human can
	4	214.8	43.0	662	14	ACL58180	Acl58180 Human col
	5	214.8	43.0	35804	14	ADZ59751	Adz59751 Human CB1
	6	214	42.8	596	5	ABA17596	Aba17596 Human ner
	7	213.4	42.7	587	6	ABN64811	Abn64811 Human can
	8	213.4	42.7	587	14	ACL57703	Acl57703 Human col
	9	213.4	42.7	9469	4	AAK79514	Aak79514 Human imm
	10	213.4	42.7	168174	6	ABT11173	Abt11173 Human 5-l
	11	213.4	42.7	168273	6	ABT11114	Abt11114 Human 5-l
c	12	212.6	42.6	3660	12	ADQ63479	Adq63479 Novel hum
	13	211.8	42.4	10252	4	AAK90931	Aak90931 Human dig
	14	211.8	42.4	10252	5	AAS31966	Aas31966 Human liv
	15	211.8	42.4	10252	6	ABN90321	Abn90321 Human liv
	16	211.8	42.4	10252	11	ADJ15234	Adj15234 Human liv
	17	211.6	42.4	127722	12	ADQ97301	Adq97301 Human can
	18	211.4	42.3	91080	6	ABX08336_16	Continuation (17 o
	19	211.4	42.3	91134	12	ADO50281_16	Continuation (17 o
	20	211.4	42.3	91138	12	ADN97989_16	Continuation (17 o
	21	211.4	42.3	91139	12	ADJ25985_16	Continuation (17 o
	22	211.4	42.3	91140	14	AEB85185_16	Continuation (17 o
c	23	211	42.2	3723	11	ADM01799	Adm01799 Human cDN
	24	211	42.2	11337	4	AAL36286	Aal36286 Human mus
	25	211	42.2	11337	8	ABX59274	Abx59274 cDNA enco
	26	211	42.2	11337	12	ADJ30024	Adj30024 Human mus
c	27	210.6	42.2	4622	4	ABK42254	Abk42254 Genomic s
c	28	210.6	42.2	4622	9	ADB60410	Adb60410 Connectiv
	29	210.6	42.2	11813	12	ADM80714	Adm80714 Human cla
	30	210.6	42.2	52216	4	AAH28355	Aah28355 Nucleotid
	31	210.6	42.2	52216	6	ABL50307	Abl50307 Human mus
c	32	210.4	42.1	10113	4	AAS32838	Aas32838 Human gen
c	33	210.4	42.1	11655	4	AAS32839	Aas32839 Human gen
c	34	210.4	42.1	30652	14	AEB94744	Aeb94744 Human eIF
	35	210	42.0	79544	12	ADQ97764	Adq97764 Human can
c	36	210	42.0	110000	13	ADS99457_0	Ads99457 Human kin
	37	209.8	42.0	2085	13	ADR07789	Adr07789 Full leng
	38	209.8	42.0	6867	5	ABA14921	Aba14921 Human ner
	39	209.8	42.0	7270	5	ABA14924	Aba14924 Human ner
	40	209.8	42.0	110000	11	ACN45090_1	Continuation (2 of
	41	209.8	42.0	110000	11	ACN45090_2	Continuation (3 of
	42	209.4	41.9	2183	4	AAH17953	Aah17953 Human cDN
c	43	209.2	41.9	27204	11	ACN44866	Acn44866 Human gen
c	44	209	41.8	1990	4	ABK42788	Abk42788 Genomic s
	45	209	41.8	1990	4	AAK87346	Aak87346 Human imm

ALIGNMENTS

RESULT 1
ADI35082
ID ADI35082 standard; DNA; 12174 BP.
XX
AC ADI35082;
XX
DT 22-APR-2004 (first entry)
XX
DE Human PLA2G1B nucleotide sequence.
XX
KW PLA2G1B ; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

```
CC   site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC   subject, where the presence of the polymorphic variation indicates a
CC   predisposition to fat deposition in the subject. The polymorphic
CC   variation is a guanine at position 7328 or thymine at position 9182 of
CC   the present sequence. The method is useful for diagnosing a
CC   predisposition to fat deposition or leanness in a subject, and
CC   consequently for diagnosing a predisposition to non-insulin dependent
CC   diabetes mellitus (NIDDM) in a subject and conditions such as
CC   hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC   hypercoagulability, or microalbuminuria, which can lead to early
CC   prescription of preventive measures. The present sequence represents a
CC   human PLA2G1B nucleotide sequence.
XX
SQ   Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 499.6;  DB 12;  Length 12174;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-146;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60

Qy         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120

Qy        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180

Qy        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240

Qy        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300

Qy        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 360

Qy        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420

Qy        421 TCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480

Qy        481 ACCCACATAAGAGGGACTGA 500
              ||||||||||||||||||||
Db        481 ACCCACATAAGAGGGACTGA 500


RESULT 2
ADJ09983
ID   ADJ09983 standard; DNA; 12174 BP.
XX
AC   ADJ09983;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW   human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW   PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW   appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW   non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW   hypertension; antidiabetic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(436,c)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(839,a)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4050,a)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4689,t)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6282,a)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6358,c)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6653,t)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7256,t)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7300,a)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7301,c)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7328,g)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(8062,c)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(9182,t)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(11649,c)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphisms"
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   P-PSDB; ADJ09984.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
```

```
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ   Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 499.6;  DB 12;  Length 12174;
  Best Local Similarity   99.8%;  Pred. No. 3.3e-146;
  Matches  499;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTACCTCGACCTTTGTGCCAGGTTCTTAGCATATGGGACCTGGGATGGAGTTAGCGC 60

Qy         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120

Qy        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180

Qy        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240

Qy        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300

Qy        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAAGGGTAATTAATAAC 360

Qy        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCA 420

Qy        421 TCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCTTCAGGCCTAAATCTAGGAGCAGTGCCTGGTCCTCTGTCTTGTTATGACCCCAAGGA 480

Qy        481 ACCCACATAAGAGGGACTGA 500
              ||||||||||||||||||||
Db        481 ACCCACATAAGAGGGACTGA 500


RESULT 3
ABN61073
ID   ABN61073 standard; cDNA; 662 BP.
XX
AC   ABN61073;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human cancer related polynucleotide SEQ ID NO 1040.
XX
KW   Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW   gene therapy; cancer; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200214500-A2.
XX
PD   21-FEB-2002.
XX
PF   16-AUG-2001; 2001WO-US025840.
XX
PR   16-AUG-2000; 2000US-0226326P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Escobedo J,  Garcia PD,  Sudduth-Klinger J,  Reinhard C,  Randazzo F;
PI   Lamson G,  Scott EM,  Zhang G,  Kassam A,  Pot D,  Labat I;
XX
DR   WPI; 2002-241905/29.
XX
PT   New nucleic acid for producing a polypeptide, detecting differentially
PT   expressed genes correlated with a cancerous state of a mammalian cell,
PT   and inhibiting tumor growth.
XX
PS   Claim 1; SEQ ID NO 1040; 883pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC   with cytostatic activity. The polynucleotide is used to produce a
CC   polypeptide, to detect differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell and to inhibit tumour growth. The
CC   polynucleotide is used as a probe in mapping and tissue profiling. The
CC   encoded polypeptide and antibodies to the polypeptide can also be used
CC   for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC   gene therapy. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 662 BP; 202 A; 136 C; 174 G; 150 T; 0 U; 0 Other;

  Query Match             43.0%;  Score 214.8;  DB 6;  Length 662;
  Best Local Similarity   83.8%;  Pred. No. 5.4e-57;
  Matches  243;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         63 AGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGG 122
              ||| || |||     ||| ||||||  ||||||||||   ||||| ||||||||||||||
Db        184 AGTAAAGAGTTTTGGATTGGCCAGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGG 243

Qy        123 GAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGG 182
              ||| |||||   ||||||||||||||  ||||||||  |||||||| ||||||||||||
Db        244 GAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGG 303

Qy        183 CAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
              ||||  || ||||||||||||||||||||||||||||||||||||||||  ||||  ||
Db        304 TGAAACCCTGTCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGCGCCTGTA 363

Qy        243 GTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTG 302
              | ||||||||| | || ||||| |||||| ||||||||||||||||||||||||  ||
Db        364 GCCCCAGCTACTCTGGCGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGCTG 423

Qy        303 CAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||  |||| ||||||||||| ||||||||||  || ||  || |
Db        424 CAGTGAGCTGAGATCGCACCACTGCATTCCAGCCTGGGCAACAAAAGTGA 473


RESULT 4
ACL58180
ID   ACL58180 standard; cDNA; 662 BP.
XX
AC   ACL58180;
XX
DT   24-MAR-2005  (first entry)
XX
DE   Human colon cancer differentially expressed polynucleotide, SEQ ID:4315.
XX
KW   Differential expression; diagnosis; therapy; drug screening; cancer;
KW   neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO2005000087-A2.
```

XX
PD 06-JAN-2005.
XX
PF 13-MAY-2004; 2004WO-US015421.
XX
PR 03-JUN-2003; 2003US-0475872P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Randazzo F, Moler E, Escobedo J, Garcia PD;
XX
DR WPI; 2005-075421/08.
XX
PT New isolated polynucleotides, which are differentially expressed in colon
PT cancer cell, useful for treating cancer, e.g. colon cancer, breast
PT cancer, or pancreatic cancer.
XX
PS Claim 1; SEQ ID NO 4315; 97pp; English.
XX
CC The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which
CC are differentially expressed in colon cancer cells. The invention also
CC relates to vectors and host cells comprising a differentially expressed
CC polynucleotide of the invention; a method for detecting a cancerous cell
CC by detection of a gene product of the polynucleotides; a method for
CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product
CC of the polynucleotides; a method of treating an individual with cancer by
CC administration of a modulator of a gene product of the polynucleotides;
CC and an isolated antibody that specifically binds to a polypeptide encoded
CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,
CC antibodies, and methods are useful for the detection of cancerous cells;
CC for the diagnosis, prognosis and management of cancer; for the
CC identification of agents that modulate the phenotype of cancerous cells;
CC for the identification of therapeutic targets for cancer chemotherapy;
CC and for the treatment of cancer, especially colon cancer and metastasized
CC colon cancer, but also breast or pancreatic cancer. The polynucleotides
CC are also useful as a source of probes or primers for use in diagnostic
CC methods. The differentially expressed polynucleotides or their encoded
CC proteins can additionally be used as vaccines to modulate primary immune
CC responses for the prevention or treatment of cancer. The present sequence
CC represents a specifically claimed polynucleotide which is differentially
CC expressed in colon cancer. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 662 BP; 202 A; 136 C; 174 G; 150 T; 0 U; 0 Other;

Query Match 43.0%; Score 214.8; DB 14; Length 662;
Best Local Similarity 83.8%; Pred. No. 5.4e-57;
Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy         63 AGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGG 122
              ||| || |||     ||| ||||||  ||||||||||   ||||| ||||||||||||||
Db        184 AGTAAAGAGTTTTGGATTGGCCAGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGG 243

Qy        123 GAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGG 182
              ||| |||||   ||||||||||||||  ||||||||| ||||||||| ||||||||||||
Db        244 GAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGG 303

Qy        183 CAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
                |||| || ||||||||||||||||||||||||||||||||||||||||||  |||| ||
Db        304 TGAAACCCTGTCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGCGCCTGTA 363

Qy        243 GTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTG 302
              | ||||||||| | || ||||| ||||| |||||| ||||||||||||||||||||| ||
Db        364 GCCCCAGCTACTCTGGCGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGCTG 423

Qy        303 CAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||  |||| |||||||||||  ||||||||||  || ||  || |
Db        424 CAGTGAGCTGAGATCGCACCACTGCATTCCAGCCTGGGCAACAAAAGTGA 473
```

RESULT 5

ADZ59751
ID ADZ59751 standard; DNA; 35804 BP.
XX
AC ADZ59751;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human CB1 cannabinoid receptor interacting protein 1b gene SEQ ID NO:3.
XX
KW CB1 cannabinoid receptor interacting protein 1; antiemetic; anabolic;
KW analgesic; inotropic; ophthalmological; gastrointestinal-gen.; nootropic;
KW neuroprotective; analgesia; multiple sclerosis; nausea; drug-induced;
KW cancer; chemotherapy; retinoblastoma; chromosome-2; gene; ds.
XX
OS Homo sapiens.
XX
PN US2005095674-A1.
XX
PD 05-MAY-2005.
XX
PF 23-JUL-2004; 2004US-00898406.
XX
PR 23-JUL-2003; 2003US-0489542P.
PR 27-FEB-2004; 2004US-0548697P.
XX
PA (LEWI/) LEWIS D L.
PA (BHAR/) BHARTUR S.
PA (WALL/) WALLIS K.
PA (NIEH/) NIEHAUS J.
XX
PI Lewis DL, Bhartur S, Wallis K, Niehaus J;
XX
DR WPI; 2005-332074/34.
XX
PT Novel isolated cannabinoid receptor nucleic acid, useful for modulating
PT activity of CB1 cannabinoid receptor such as appetite stimulation,
PT analgesia, euphoria, decreased tremor or spasticity associated with
PT multiple sclerosis.
XX
PS Claim 1; SEQ ID NO 3; 77pp; English.
XX
CC The invention relates to an isolated CB1 cannabinoid receptor interacting
CC protein 1 (CRIP1) nucleic acid (I). Also described: (1) a vector
CC comprising (I); (2) a composition (II) comprising (I) and a carrier; (3)
CC modulating the activity of a CB1 cannabinoid receptor comprising
CC administering (II) to the CB1 cannabinoid receptor; (4) an antibody that
CC specifically binds to a CRIP1 polypeptide; and (5) screening for a
CC modulator of CRIP1 activity on a CB1 cannabinoid receptor. (I) and (II)
CC are useful for modulating the activity of a CB1 cannabinoid receptor,
CC which involves administering (I) or (II) to the CB1 cannabinoid receptor
CC in a cell, which is a human cell of a patient. The modulation of CB1
CC cannabinoid receptor in the patient results in one or more effects chosen
CC from appetite stimulation, analgesia, euphoria, decreased tremor or
CC spasticity associated with multiple sclerosis, attenuation of nausea and
CC vomiting in cancer, chemotherapy, reduction in intraocular pressure,
CC decreased intestinal motility, and attenuation of aversive memories,
CC appetite suppression, decreased sedation, decreased alterations in
CC cognition and memory, and decreased mood alterations. (II) is useful for
CC modulating the tonic activity of the CB1 cannabinoid receptor, which
CC involves administering (II) to the CB1 cannabinoid receptor. Human CRIP1a
CC and CRIP1b map to chromosome 2, and are retinoblastoma genes. The present
CC sequence represents the human CRIP1b genomic DNA sequence.
XX
SQ Sequence 35804 BP; 11398 A; 7034 C; 7287 G; 10085 T; 0 U; 0 Other;

Query Match 43.0%; Score 214.8; DB 14; Length 35804;
Best Local Similarity 83.8%; Pred. No. 2.9e-56;
Matches 243; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy         63 AGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGG 122
              ||| || |||     ||| ||||||  ||||||||||   ||||| ||||||||||||||
Db      25857 AGTAAAGAGTTTTGGATTGGCCAGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGG 25916
```

```
Qy        123 GAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGG 182
              ||| |||||   ||||||||||||||  |||||||| |||||||| ||||||||||||
Db      25917 GAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGG 25976

Qy        183 CAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATA 242
              |||| || ||||||||||||||||||||||||||||||||||||||||  |||| ||
Db      25977 TGAAACCCTGTCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACGCGCCTGTA 26036

Qy        243 GTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTG 302
              | ||||||||| | || ||||| |||| |||||| |||||||||||||||||||| ||
Db      26037 GCCCCAGCTACTCTGGCGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGTGGAGGCTG 26096

Qy        303 CAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||  |||| |||||||||| |||||||||  || ||  || |
Db      26097 CAGTGAGCTGAGATCGCACCACTGCATTCCAGCCTGGGCAACAAAAGTGA 26146
```

RESULT 6
ABA17596
ID ABA17596 standard; DNA; 596 BP.
XX
AC ABA17596;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9927.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
PS Disclosure; SEQ ID NO 9927; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 596 BP; 177 A; 123 C; 167 G; 129 T; 0 U; 0 Other;

Query Match 42.8%; Score 214; DB 5; Length 596;
Best Local Similarity 79.6%; Pred. No. 9.3e-57;
Matches 253; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         35 TATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTG 94
              ||||  |||  |  |||  |   |   |||| | |||       |||||||||  |||
Db        254 TATGATGCCTCCAAATGAGCCAAGACAGTCTTAAAAATAAGATTGAGGCCAGGTGCAGTG 313

Qy         95 GCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAG 154
              ||||||| ||||| |||||||||||||||| ||||| |||||||||||| ||||  | |
Db        314 GCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACCTGAGGTCGG 373

Qy        155 GAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAAAAAT 214
              ||||| |||||||||||| |||||||||  |||| || |||||| ||||||||||||||||
Db        374 GAGTTCGAGACCAGCCTGACCAACATGGAGAAACCCTGTCTCTACTAAAAATACAAAAAT 433

Qy        215 TAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGA 274
              ||||  ||  ||||||| | ||||| ||||||||||||| | |||||||| ||||| |||
Db        434 TAGCTGGGCATGGTGGCGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 493

Qy        275 ATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAG 334
              ||  |||||||||||||||||||||||||| ||||| || ||| ||||||||||||||||
Db        494 ATTGCTTGAACCTGGGAGGTGGAGGTTGCGGTGAGTCAGGATCGCACCACTGCACTCCAG 553

Qy        335 CCTGGAAAAAAAAGGGTAA 352
              |||||  ||  || | ||
Db        554 CCTGGGGAACGAGAGCAA 571
```

RESULT 7
ABN64811
ID ABN64811 standard; cDNA; 587 BP.
XX
AC ABN64811;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 4778.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US025840.
XX
PR 16-AUG-2000; 2000US-0226326P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth.
XX
PS Claim 1; SEQ ID NO 4778; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 587 BP; 149 A; 131 C; 163 G; 144 T; 0 U; 0 Other;

Query Match 42.7%; Score 213.4; DB 6; Length 587;

```
  Best Local Similarity   78.3%;  Pred. No. 1.4e-56;
  Matches  256;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCA 85
              || |||  || || | ||   ||| |  |||    | |   || || ||     | ||
Db        255 TTTTTATGATGTGAGTCCCAAGATAGTTTTACAAATGATAAAAGAGGAAGAAGCTGGCTG 314

Qy         86 GGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACT 145
              || || |||||||| | ||||| |||||||||||||||||| | |||  ||||||||||
Db        315 GGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACC 374

Qy        146 TGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAA 205
              ||||  |||||||||||||||||| |||||||| |||| |||| |  |||||| ||||||
Db        375 TGAGGTCAGGAGTTTGAGACCAGCTTGGCCAACGTGGCGAAACCCCGTCTCTACTAAAAA 434

Qy        206 TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGG 265
              ||||||||||||| || ||||||||    |||| || |||||||||| ||||||||||
Db        435 TACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGA 494

Qy        266 GGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACT 325
              ||||| |||||  |||||||| |||||| |||||||||||||||||||||||| ||| |
Db        495 GGCAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGCGCCATT 554

Qy        326 GCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||||||||  || ||| || |
Db        555 GCACTCCAGCCTGGGCAACAAGAGTGA 581


RESULT 8
ACL57703
ID   ACL57703 standard; cDNA; 587 BP.
XX
AC   ACL57703;
XX
DT   24-MAR-2005  (first entry)
XX
DE   Human colon cancer differentially expressed polynucleotide, SEQ ID:3838.
XX
KW   Differential expression; diagnosis; therapy; drug screening; cancer;
KW   neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO2005000087-A2.
XX
PD   06-JAN-2005.
XX
PF   13-MAY-2004; 2004WO-US015421.
XX
PR   03-JUN-2003; 2003US-0475872P.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Randazzo F,  Moler E,  Escobedo J,  Garcia PD;
XX
DR   WPI; 2005-075421/08.
XX
PT   New isolated polynucleotides, which are differentially expressed in colon
PT   cancer cell, useful for treating cancer, e.g. colon cancer, breast
PT   cancer, or pancreatic cancer.
XX
PS   Claim 1; SEQ ID NO 3838; 97pp; English.
XX
CC   The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which
CC   are differentially expressed in colon cancer cells. The invention also
CC   relates to vectors and host cells comprising a differentially expressed
CC   polynucleotide of the invention; a method for detecting a cancerous cell
CC   by detection of a gene product of the polynucleotides; a method for
CC   inhibiting a cancerous phenotype of a cell by inhibiting a gene product
CC   of the polynucleotides; a method of treating an individual with cancer by
CC   administration of a modulator of a gene product of the polynucleotides;
CC   and an isolated antibody that specifically binds to a polypeptide encoded
CC   by one of the 9672 polynucleotides. The polynucleotides, polypeptides,
CC   antibodies, and methods are useful for the detection of cancerous cells;
CC   for the diagnosis, prognosis and management of cancer; for the
CC   identification of agents that modulate the phenotype of cancerous cells;
CC   for the identification of therapeutic targets for cancer chemotherapy;
CC   and for the treatment of cancer, especially colon cancer and metastasized
CC   colon cancer, but also breast or pancreatic cancer. The polynucleotides
CC   are also useful as a source of probes or primers for use in diagnostic
CC   methods. The differentially expressed polynucleotides or their encoded
CC   proteins can additionally be used as vaccines to modulate primary immune
CC   responses for the prevention or treatment of cancer. The present sequence
CC   represents a specifically claimed polynucleotide which is differentially
CC   expressed in colon cancer. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 587 BP; 149 A; 131 C; 163 G; 144 T; 0 U; 0 Other;

  Query Match           42.7%;  Score 213.4;  DB 14;  Length 587;
  Best Local Similarity   78.3%;  Pred. No. 1.4e-56;
  Matches  256;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCA 85
              || |||  || || | ||   ||| |  |||    | |   || || ||     | ||
Db        255 TTTTTATGATGTGAGTCCCAAGATAGTTTTACAAATGATAAAAGAGGAAGAAGCTGGCTG 314

Qy         86 GGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACT 145
              || || |||||||| | ||||| |||||||||||||||||| | |||  ||||||||||
Db        315 GGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACC 374

Qy        146 TGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAA 205
              ||||  |||||||||||||||||| |||||||| |||| |||| |  |||||| ||||||
Db        375 TGAGGTCAGGAGTTTGAGACCAGCTTGGCCAACGTGGCGAAACCCCGTCTCTACTAAAAA 434

Qy        206 TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGG 265
              ||||||||||||| || ||||||||    |||| || |||||||||| ||||||||||
Db        435 TACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGA 494

Qy        266 GGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACT 325
              ||||| |||||  |||||||| |||||| |||||||||||||||||||||||| ||| |
Db        495 GGCAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGCGCCATT 554

Qy        326 GCACTCCAGCCTGGAAAAAAAGGGTAA 352
              ||||||||||||||  || ||| || |
Db        555 GCACTCCAGCCTGGGCAACAAGAGTGA 581


RESULT 9
AAK79514
ID   AAK79514 standard; DNA; 9469 BP.
XX
AC   AAK79514;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34326.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
```

PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 34326; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic

CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 9469 BP; 2940 A; 2203 C; 1941 G; 2385 T; 0 U; 0 Other;

Query Match 42.7%; Score 213.4; DB 4; Length 9469;
Best Local Similarity 84.0%; Pred. No. 4.6e-56;
Matches 241; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy        66 TAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAG 125
             | |  | || |  |  |||||| | |||||||||||| ||||| ||||||||||||||||||
Db      1143 TGACGGAAAATGGTCGGCCAGGCGAGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAG 1202

Qy       126 ACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAA 185
             || || |||| ||||||| ||||  ||||||||  |||||||||||||||||||||  |
Db      1203 GCCAAGGTGGGAGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGA 1262

Qy       186 AACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTC 245
             |||  |  |||||| |||||||||||||||||||  ||  |||||||||  ||||  || ||
Db      1263 AACCCCGTCTCTACTAAAAATACAAAAATTAGCAGGGCATGGTGGCACGCGCCTGTAATC 1322

Qy       246 CCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAG 305
             |||||||| |||||||||| |||||||||||||||||||||| |||||| ||||||||||
Db      1323 CCAGCTACTCAGGAGGCTGAGGCAGAAGAATCACTTGAACCCGGGAGGCGGAGGTTGCAG 1382

Qy       306 TGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
             |||||||||| |||||||||||||||||||| || || ||| | ||
Db      1383 TGAGCCAAGATGGCACCACTGCACTCCAGCCTAGACAACAAGAGCAA 1429
```

RESULT 10
ABT11173
ID ABT11173 standard; DNA; 168174 BP.
XX
AC ABT11173;
XX
DT 05-DEC-2002 (first entry)
XX
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 63.
XX
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200262825-A2.
XX
PD 15-AUG-2002.
XX
PF 07-FEB-2002; 2002WO-US003546.
XX
PR 08-FEB-2001; 2001US-0267515P.
PR 21-AUG-2001; 2001US-0314248P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Barnes G, Meyer J;
XX
DR WPI; 2002-627522/67.
XX
PT New isolated nucleic acid molecule with an allelic variant of a
PT polymorphic region of an 5-LO gene, useful for diagnosing and/or
PT prognosticating disorders associated with an aberrant inflammatory
PT response such as asthma.
XX
PS Disclosure; Fig 4; 290pp; English.
XX
CC The invention relates to an isolated human nucleic acid molecule
CC comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
CC (5-LO) gene, where the allelic variant comprises one or more nucleotide
CC selected from any of 3, 20 or 21 base pair sequences, given in the
CC specification, or their complement. The compositions and methods of the
CC present invention are useful for diagnosing and/or prognosing disorders
CC associated with an aberrant inflammatory response such as asthma,
CC bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
CC disease, atopic and contact dermatitis. The nucleic acid molecules can
CC also be useful for identifying an individual amongst other individuals
CC from the same species for use in forensic medicine and paternity testing.
CC This polynucleotide sequence represents DNA relating to the human 5-
CC lipoxygenase (5-LO) gene of the invention
XX
SQ Sequence 168174 BP; 46808 A; 36442 C; 36942 G; 46474 T; 0 U; 1508 Other;

Query Match 42.7%; Score 213.4; DB 6; Length 168174;
Best Local Similarity 78.3%; Pred. No. 1.5e-55;
Matches 256; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy        26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCA 85
             || |||  || || | ||   ||| |  |||     | |   || || ||      | ||
Db     26811 TTTTTATGATGTGAGTCCCAAGATAGTTTTACAAATGATAAAAGAGGAAGAAGCTGGCTG 26870

Qy        86 GGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACT 145
             || || |||||||| | ||||| |||||||||||||||| | |||  ||||||||||
Db     26871 GGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACC 26930

Qy       146 TGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAA 205
             ||||  |||||||||||||||||| |||||||| |||| |||| |  |||||| ||||||
Db     26931 TGAGGTCAGGAGTTTGAGACCAGCTTGGCCAACGTGGCGAAACCCCGTCTCTACTAAAAA 26990

Qy       206 TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGG 265
             |||||||||||||| || ||||||||    |||| || |||||||||| ||||||||||
Db     26991 TACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGA 27050

Qy       266 GGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACT 325
             ||||| |||||  |||||||| |||||| ||||||||||||||||||||||||| ||| |
Db     27051 GGCAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGCGCCATT 27110

Qy       326 GCACTCCAGCCTGGAAAAAAAGGGTAA 352
             |||||||||||||  || ||| || |
Db     27111 GCACTCCAGCCTGGGCAACAAGAGTGA 27137
```

RESULT 11
ABT11114
ID ABT11114 standard; DNA; 168273 BP.
XX
AC ABT11114;
XX
DT 05-DEC-2002 (first entry)
XX
DE Human 5-lipoxygenase gene related DNA sequence SEQ ID No 2.
XX
KW Human; polymorphic region; 5-lipoxygenase; 5-LO gene; asthma; bronchitis;
KW sinusitis; ulcerative colitis; nephritis; amyloidosis; sarcoidosis;
KW rheumatoid arthritis; scleroderma; lupus; non-allergic rhinitis;
KW polymyositis; Reiter's syndrome; psoriasis; pelvic inflammatory disease;
KW atopic; contact dermatitis; forensic medicine; paternity testing; enzyme;

```
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200262825-A2.
XX
PD   15-AUG-2002.
XX
PF   07-FEB-2002; 2002WO-US003546.
XX
PR   08-FEB-2001; 2001US-0267515P.
PR   21-AUG-2001; 2001US-0314248P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Barnes G,  Meyer J;
XX
DR   WPI; 2002-627522/67.
XX
PT   New isolated nucleic acid molecule with an allelic variant of a
PT   polymorphic region of an 5-LO gene, useful for diagnosing and/or
PT   prognosticating disorders associated with an aberrant inflammatory
PT   response such as asthma.
XX
PS   Disclosure; Fig 2; 290pp; English.
XX
CC   The invention relates to an isolated human nucleic acid molecule
CC   comprising an allelic variant of a polymorphic region of a 5-lipoxygenase
CC   (5-LO) gene, where the allelic variant comprises one or more nucleotide
CC   selected from any of 3, 20 or 21 base pair sequences, given in the
CC   specification, or their complement. The compositions and methods of the
CC   present invention are useful for diagnosing and/or prognosing disorders
CC   associated with an aberrant inflammatory response such as asthma,
CC   bronchitis, sinusitis, ulcerative colitis, nephritis, amyloidosis,
CC   rheumatoid arthritis, sarcoidosis, scleroderma, lupus, non-allergic
CC   rhinitis, polymyositis, Reiter's syndrome, psoriasis, pelvic inflammatory
CC   disease, atopic and contact dermatitis. The nucleic acid molecules can
CC   also be useful for identifying an individual amongst other individuals
CC   from the same species for use in forensic medicine and paternity testing.
CC   This polynucleotide sequence represents DNA relating to the human 5-
CC   lipoxygenase (5-LO) gene of the invention
XX
SQ   Sequence 168273 BP; 46834 A; 36467 C; 36966 G; 46498 T; 0 U; 1508 Other;

  Query Match             42.7%;  Score 213.4;  DB 6;  Length 168273;
  Best Local Similarity   78.3%;  Pred. No. 1.5e-55;
  Matches  256;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         26 TTCTTAGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCA 85
Db      26861 TTTTTATGATGTGAGTCCCAAGATAGTTTTACAAATGATAAAAGAGGAAGAAGCTGGCTG 26920

Qy         86 GGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACT 145
Db      26921 GGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACC 26980

Qy        146 TGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAA 205
Db      26981 TGAGGTCAGGAGTTTGAGACCAGCTTGGCCAACGTGGCGAAACCCCGTCTCTACTAAAAA 27040

Qy        206 TACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGG 265
Db      27041 TACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGA 27100

Qy        266 GGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACT 325
Db      27101 GGCAGGAGAATTGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCAAGATTGCGCCATT 27160

Qy        326 GCACTCCAGCCTGGAAAAAAAGGGTAA 352
Db      27161 GCACTCCAGCCTGGGCAACAAGAGTGA 27187
```

```
RESULT 12
ADQ63479/c
ID   ADQ63479 standard; cDNA; 3660 BP.
XX
AC   ADQ63479;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human cDNA sequence #640.
XX
KW   ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW   cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Nagai K,  Irie R;
XX
DR   WPI; 2004-535376/52.
DR   P-PSDB; ADQ65667.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 640; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide
CC   sequence of the invention.
XX
SQ   Sequence 3660 BP; 887 A; 1049 C; 837 G; 887 T; 0 U; 0 Other;

  Query Match             42.6%;  Score 212.6;  DB 12;  Length 3660;
  Best Local Similarity   78.6%;  Pred. No. 5.5e-56;
  Matches  254;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy         37 TGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGC 96
Db       1862 TGTGGCCTGAGCTGGCATCAGACAGGCTTTAAGAGGAGAATGTCAGCTGGGTGCGGTGGC 1803

Qy         97 TCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGA 156
Db       1802 TCACACCTGTAATTCTAGCACTTTGGGAGGCCGAGGTGGGTGGATCCCTTGAGGTCAGGA 1743

Qy        157 GTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAAAAATTA 216
Db       1742 GTTCGAGACCAGCCTGGCCAACATGGCAAAACCCCATCTCTATGAAAAATAGAAAAATTA 1683

Qy        217 GCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAAT 276
Db       1682 GCCAGGCGTGGTGGCAGGCGCCTGTAATTCCAGCTACTCAGGAGGCTGAGGCACAAGAAT 1623

Qy        277 CACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCC 336
```

```
Db       1622 CACTTGAACCCGGGAGGCAGAGGTTGCAATGAGCCAAGACTGCGCCACTGCACTCCAGCC 1563

Qy        337 TGGAAAAAAAGGGTAATTAATAA 359
              |||   | |  |  |  |  |||
Db       1562 TGGGCGACAGAGTGAGGTCCTAA 1540
```

RESULT 13
AAK90931
ID AAK90931 standard; DNA; 10252 BP.
XX
AC AAK90931;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4507.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001324.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

```
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases.
XX
PS   Disclosure; SEQ ID NO 4507; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention
XX
SQ   Sequence 10252 BP; 2689 A; 2608 C; 2537 G; 2418 T; 0 U; 0 Other;

  Query Match             42.4%;  Score 211.8;  DB 4;  Length 10252;
  Best Local Similarity   80.1%;  Pred. No. 1.5e-55;
  Matches  249;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          64 GTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGG 123
               ||||||| |       |||||| || |||||||||| ||||| ||||||||||||||||
Db        8374 GTTAATAAAGAAGGTGGGGCCAGGAGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 8433

Qy         124 AGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGC 183
               || | |||  |||||||||| ||||  |||||||| ||||||||||||||||||||| |
Db        8434 AGTCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGAC 8493

Qy         184 AAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAG 243
               ||||| | ||||||| |||||||||||||||||||  |||||||||||| || ||||||
Db        8494 AAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCAGTTTCCTATAA 8553

Qy         244 TCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGC 303
               || ||||||| |||||||||| ||||| |||||| ||||||||  |||||  ||||||||
Db        8554 TCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCCGAGGTTGC 8613

Qy         304 AGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTT 363
               |||||||||||||||| |||||||||||||||||||  || |  | | |   | ||
Db        8614 AGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGATTCCATCTCA 8673

Qy         364 ACTTGCAACCA 374
               |     ||| |
Db        8674 AAACAAAACAA 8684


RESULT 14
AAS31966
```

```
ID   AAS31966 standard; DNA; 10252 BP.
XX
AC   AAS31966;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human liver associated genomic DNA #140.
XX
KW   Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155355-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001351.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
```

PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457728/49.
XX
PT Isolated nucleic acid molecule encoding a human liver related protein is
PT used in preventing, treating or ameliorating disorders of the liver
PT particularly cancer of the liver.
XX
PS Claim 1; SEQ ID NO 442; 526pp; English.
XX
CC Sequences AAS31827-AAS32182 represent genomic DNA molecules, which encode
CC the liver associated polypeptides of the invention. Liver associated
CC polypeptides and their associated polynucleotides are useful in the
CC diagnosis, treatment and prevention of various types of disorders in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a liver associated polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 10252 BP; 2689 A; 2608 C; 2537 G; 2418 T; 0 U; 0 Other;

Query Match 42.4%; Score 211.8; DB 5; Length 10252;
Best Local Similarity 80.1%; Pred. No. 1.5e-55;
Matches 249; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 64 GTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGG 123
Db 8374 GTTAATAAAGAAGGTGGGGCCAGGAGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 8433

Qy 124 AGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGC 183
Db 8434 AGTCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGAC 8493

Qy 184 AAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAG 243
Db 8494 AAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCAGTTTCCTATAA 8553

Qy 244 TCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGC 303

```
Db       8554 TCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCCGAGGTTGC 8613

Qy        304 AGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTT 363
              |||||||||||||||| ||||||||||||||||||  || |  |  |  |    | ||
Db       8614 AGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGATTCCATCTCA 8673

Qy        364 ACTTGCAACCA 374
              |     ||| |
Db       8674 AAACAAAACAA 8684


RESULT 15
ABN90321
ID   ABN90321 standard; DNA; 10252 BP.
XX
AC   ABN90321;
XX
DT   24-JUL-2002  (first entry)
XX
DE   Human liver antigen HLDQN90 genomic sequence, SEQ ID NO:442.
XX
KW   Human; liver antigen; liver disorder; hepatic disorder; infection;
KW   hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;
KW   cirrhosis; granulomatous hepatitis; toxin damage; drug damage;
KW   autoimmune disease; Wilson's disease; primary biliary cirrhosis;
KW   neoplastic disorder; cancer; tumour; portal hypertension;
KW   gastrointestinal disorder; hepatitis; drug screening; gene therapy;
KW   chromosome mapping; forensic analysis; antibody preparation;
KW   hepatotropic; cytostatic; antiinflammatory; virucide; antibacterial;
KW   fungicide; parasiticide; antidote; immunosuppressive; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2002042096-A1.
XX
PD   11-APR-2002.
XX
PF   17-JAN-2001; 2001US-00764887.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-381944/41.
XX
PT   New nucleic acid encoding human liver antigens, useful for diagnosis,
PT   treatment and prevention of e.g. hepatitis and hepatic cancer, also
PT   related polypeptides and antibodies.
XX
PS   Disclosure; SEQ ID NO 442; 181pp; English.
XX
CC   The invention relates to 145 novel human liver antigens (ABP40831-
CC   ABP40975) and to cDNAs encoding them (ABN90036-ABN90180), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human liver antigen
CC   polynucleotides, antibodies against human liver antigens, and the use of
CC   liver antigen polynucleotides and polypeptides in diagnosing, treating,
CC   prognosing or preventing various disorders of the liver. Such conditions
CC   include viral infections (e.g., cytomegalovirus, Epstein-Barr virus,
CC   hepatitis A virus, hepatitis B virus and hepatitis C virus), parasitic
CC   infections (e.g., Clonorchis sinensis, Echinococcus granulosus and
CC   Entamoeba histolytica), and also bacterial and fungal infections. Other
CC   disorders that may be treated include inflammatory conditions (e.g.,
CC   cirrhosis and granulomatous hepatitis), damage caused by drugs or toxins,
CC   autoimmune diseases (e.g. Wilson's disease, primary biliary cirrhosis),
CC   neoplastic disorders (e.g., adenomas, haemangiomas and hepatocellular
CC   carcinoma), portal hypertension, or gastrointestinal disorders (e.g.,
CC   peptic ulcers, gastritis and peritoneal diseases). Liver antigen
CC   polypeptides and polynucleotides may also be used in screening for
CC   compounds which modulate liver antigen expression or activity. The
CC   polynucleotides may further be used for gene therapy, chromosome mapping,
CC   in the identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as molecular weight markers or to prepare
CC   antibodies useful in disease diagnosis, drug targeting and phenotyping.
CC   Sequences ABN90182-ABN90537 represent human liver antigen genomic
CC   sequences. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from the USPTO at seqdata.uspto.gov/sequence/
XX
SQ   Sequence 10252 BP; 2689 A; 2608 C; 2537 G; 2418 T; 0 U; 0 Other;

  Query Match             42.4%;  Score 211.8;  DB 6;  Length 10252;
  Best Local Similarity   80.1%;  Pred. No. 1.5e-55;
  Matches  249;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         64 GTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGG 123
              |||||||   |      ||||||  || ||||||||||  |||||  ||||||||||||||||
Db       8374 GTTAATAAAGAAGGTGGGGCCAGGAGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 8433

Qy        124 AGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGC 183
              || | |||  |||||||||||  ||||  ||||||||  |||||||||||||||||||||| |
Db       8434 AGTCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGAC 8493
```

```
Qy        184 AAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAG 243
              ||||| | ||||||| |||||||||||||||||||  |||||||||||| || ||||||
Db       8494 AAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCAGTTTCCTATAA 8553

Qy        244 TCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGC 303
              || ||||||| |||||||||| ||||| |||||| ||||||||  |||||  ||||||||
Db       8554 TCTCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCCGAGGTTGC 8613

Qy        304 AGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTT 363
              ||||||||||||||| |||||||||||||||||  || |  |  |  |   | ||
Db       8614 AGTGAGCCAAGATTGCGCCACTGCACTCCAGCCTGGGCAACAGAGCAAGATTCCATCTCA 8673

Qy        364 ACTTGCAACCA 374
              |     ||| |
Db       8674 AAACAAAACAA 8684
```

Search completed: February 9, 2006, 13:50:51
Job time : 346.271 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 2358.93 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 499.6
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	212.4	42.5	917	5	BU153359	BU153359 AGENCOURT
	2	210.8	42.2	687	11	CR958221	CR958221 Homo sapi
c	3	210.4	42.1	414	1	AW511559	AW511559 xu58b07.x
	4	209.8	42.0	719	9	AQ423663	AQ423663 CITBI-E1-
	5	209.6	42.0	683	10	AG119835	AG119835 Pan trogl
c	6	209.4	41.9	649	9	B71494	B71494 RPCI11-8N9.
	7	209.2	41.9	599	9	AQ586662	AQ586662 RPCI-11-4
c	8	208.8	41.8	799	3	BM679026	BM679026 UI-E-EO0-
	9	208.4	41.7	455	9	AQ037381	AQ037381 CIT-HSP-2
	10	206.8	41.4	374	8	F35684	F35684 HSPD32531 H
c	11	206.6	41.4	573	5	BX470494	BX470494 DKFZp686J
c	12	206.6	41.4	600	3	BM790237	BM790237 K-EST0070
c	13	206.6	41.4	657	3	BM788692	BM788692 K-EST0067
	14	206.6	41.4	795	11	CR960649	CR960649 Homo sapi
c	15	206.4	41.3	514	2	BE169870	BE169870 CM3-HT052
c	16	206.4	41.3	891	6	CD300637	CD300637 AGENCOURT
	17	206.2	41.3	666	9	AQ372692	AQ372692 RPCI11-14
	18	206	41.2	769	5	BU854207	BU854207 AGENCOURT
c	19	206	41.2	849	8	DN602222	DN602222 HESC3_78_
c	20	205.6	41.2	416	1	AI446478	AI446478 tj19g11.x
c	21	205.6	41.2	478	9	B67141	B67141 CIT-HSP-201
c	22	205.4	41.1	304	1	AI915081	AI915081 tq98e06.x
c	23	205.4	41.1	570	5	BX491191	BX491191 DKFZp686B
c	24	205.4	41.1	584	6	CB128054	CB128054 K-EST0177
c	25	205.4	41.1	697	9	AQ054805	AQ054805 CIT-HSP-2
	26	205.2	41.1	420	2	BE062545	BE062545 QV4-BT025
c	27	205.2	41.1	440	9	AQ111200	AQ111200 CIT-HSP-2
c	28	205.2	41.1	536	9	AQ480181	AQ480181 RPCI-11-2
	29	205.2	41.1	603	3	BP872932	BP872932 BP872932
c	30	205	41.0	437	1	AA644090	AA644090 ab62b04.s
	31	205	41.0	550	9	AQ377547	AQ377547 RPCI11-14
c	32	205	41.0	637	1	AI114494	AI114494 HA1076 Hu
c	33	205	41.0	4903	4	CR857861	CR857861 Pongo pyg
	34	204.8	41.0	431	9	AQ268195	AQ268195 RPCI11-73
c	35	204.8	41.0	546	9	AQ548438	AQ548438 RPCI-11-4
c	36	204.8	41.0	755	5	BX644757	BX644757 DKFZp781B
c	37	204.8	41.0	5938	4	HSM807608	BX647463 Homo sapi
c	38	204.6	41.0	371	5	BU752091	BU752091 UI-1-BB0-
c	39	204.6	41.0	386	8	T54464	T54464 yb06b05.s1
c	40	204.6	41.0	563	5	BU072494	BU072494 im47h01.x
c	41	204.6	41.0	833	9	BZ599987	BZ599987 WHACU62TR
	42	204.6	41.0	910	2	BF337320	BF337320 602034695
	43	204.4	40.9	681	2	BG618589	BG618589 602646014
c	44	204.2	40.9	467	2	BF940118	BF940118 nac68c10.
c	45	204.2	40.9	560	1	AL589348	AL589348 DKFZp451K

ALIGNMENTS

RESULT 1
BU153359/c
LOCUS BU153359 917 bp mRNA linear EST 03-SEP-2002
DEFINITION AGENCOURT_7934963 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6009715 5', mRNA sequence.
ACCESSION BU153359
VERSION BU153359.1 GI:22666891
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 917)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13196 row: j column: 20
High quality sequence stop: 529.
FEATURES Location/Qualifiers
source 1. .917
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6009715"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_68"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies. "
ORIGIN

Query Match 42.5%; Score 212.4; DB 5; Length 917;
Best Local Similarity 80.3%; Pred. No. 3e-26;

```
  Matches  249;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          67 AATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGA 126
               || |  ||   | | ||| ||||| |||||||||| ||||| ||||||||||||||||||
Db         575 AAAAAAAAAAAAGTGGCCGGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGG 516

Qy         127 CCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAA 186
               ||||| |||| |||||| || ||  |||||||||||||||||||||||||| ||||| ||
Db         515 CCGAGGTGGGAGGATCATTTAAGGCCAGGAGTTTGAGACCAGCCTGGCCAAAATGGCGAA 456

Qy         187 ACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCC 246
               || | |||||||| |||||||  |||||||||| || |||||||||   |||| || |||
Db         455 ACCCCATCTCTACTAAAAATCAAAAAATTAGCCGGGCGTGGTGGCAGGCGCCTGTAATCC 396

Qy         247 CAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGT 306
               ||||||| |||||||||| ||||| |||||| || |||||  ||||||||||||||||||
Db         395 CAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTGGAACCCAGGAGGTGGAGGTTGCAGT 336

Qy         307 GAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTTACT 366
               ||||||||||||||||||||||||||||||||  || |  |  |  |  ||| |
Db         335 GAGCCAAGATTGCACCACTGCACTCCAGCCTGGGCAACAGAGCGAGATTCTGTCTTAAAA 276

Qy         367 TGCAACCATA 376
               ||  |||
Db         275 ATAAAAAATA 266
```

RESULT 2
CR958221
LOCUS CR958221 687 bp DNA linear GSS 07-JUN-2005
DEFINITION Homo sapiens PAC end sequence of RZPDB737G0112D from genomic library (orig. Pieter J. de Jong library RPCI-1), genomic survey sequence.
ACCESSION CR958221
VERSION CR958221.1 GI:66955562
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 687)
AUTHORS Schwarz,F., Neubert,P., Schneider,D., Peters,M. and Korn,B.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg, Germany
COMMENT RZPD; RZPDB737G0112D; RP1-310P17;
derived from Pieter J. de Jong library RPCI-1;
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDB737G0112D
RZPDLIB; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737
http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737
http://www.rzpd.de/products/genomicset/
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available from RZPD;
Contact RZPD (customer.service@rzpd.de) for further information.
Clone distribution: http://www.rzpd.de/products/genomicset/
Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG)
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .687
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RZPDB737G0112D"
/sex="Male"
/clone_lib="RPCI-1"
/note="Vector: pCYPAC2; RPCI-1 Human Male PAC Library"

ORIGIN

```
  Query Match             42.2%;  Score 210.8;  DB 11;  Length 687;
  Best Local Similarity   83.6%;  Pred. No. 6e-26;
  Matches  239;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy          67 AATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGA 126
               ||  | |   | | ||||||| || |||||||| | ||||| ||||||||||||||||||
Db         108 AAAGGGACAACTTCAGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGG 167

Qy         127 CCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAA 186
               ||||| ||||||||||||||||  |||||||| ||||||||||||||||||||||  ||
Db         168 CCGAGGTGGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAA 227

Qy         187 ACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCC 246
               || |||||||||  ||||||||||||||||||  ||||||||||||  |||||||| |||
Db         228 ACCCTATCTCTACTAAAAATACAAAAATTAGTTGGGTGTGGTGGCAGGTGCCTATAATCC 287

Qy         247 CAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGT 306
               |||||||  |||||||| |||| ||||||||||||||||||||||||||||||||| ||||
Db         288 CAGCTACTTGGGAGGCTGAGGCAGCAGAATCACTTGAACCTGGGAGGTGGAGGTTTCAGT 347

Qy         307 GAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
               ||| | ||| ||  ||| |||| ||| ||||||  || ||| | ||
Db         348 GAGGCGAGACTGTGCCATTGCATTCCGGCCTGGGCAATAAGAGCAA 393
```

RESULT 3
AW511559/c
LOCUS AW511559 414 bp mRNA linear EST 03-MAR-2000
DEFINITION xu58b07.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2805877 3' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION AW511559
VERSION AW511559.1 GI:7149561
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 414)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1. .414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2805877"
/tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Ut1"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

ORIGIN

```
  Query Match            42.1%;  Score 210.4;  DB 1;  Length 414;
  Best Local Similarity   78.1%;  Pred. No. 8.2e-26;
  Matches  253;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         31 AGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGC 90
              | ||| || ||  | ||| |||           |||| |   |   |   | ||| |||||
Db        341 AACATTTGAGAAGTCGGAAGGAAAGTATTACATTTTAAGATCAGTTTGCTGGCCGGGTGC 282

Qy         91 GGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGA 150
              ||||||||| | ||||| ||||||||||||||||| ||||| | ||  ||||||||||
Db        281 GGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTAGGCTGATCACTTGAGG 222

Qy        151 GCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAA 210
              ||||||  |  |||||||||||||||||| | ||||| | ||||||| ||||||||||
Db        221 TCAGGAGCTCCAGACCAGCCTGGCCAACATAGTAAAACCCCATCTCTACTAAAAATACAA 162

Qy        211 AAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAG 270
              |||||||||||  ||||||  || ||||  |||||||||||| ||||||||||||||||
Db        161 AAATTAGCCAGGCATGGTGGTACATGCCGGTAGTCCCAGCTACTCAGGAGGCTGGGGCAG 102

Qy        271 AAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACT 330
              ||||||  |||||||| ||||||  |||||||||||||||||||  |   |||||||||
Db        101 GAGAATCGCTTGAACCCGGGAGGCGCAGGTTGCAGTGAGCCAAGATCGTGGCACTGCACT 42

Qy        331 CCAGCCTGGAAAAAAAGGGTAATT 354
              ||||||||  | ||||  |  | |
Db         41 GCAGCCTGGCGACAAAGTGAGACT 18
```

```
RESULT 4
AQ423663
LOCUS       AQ423663                719 bp    DNA     linear   GSS 23-MAR-1999
DEFINITION  CITBI-E1-2556C19.TF CITBI-E1 Homo sapiens genomic clone 2556C19,
            genomic survey sequence.
ACCESSION   AQ423663
VERSION     AQ423663.1  GI:4481387
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2556C19.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2556C19"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN
```

```
  Query Match            42.0%;  Score 209.8;  DB 9;  Length 719;
  Best Local Similarity   83.5%;  Pred. No. 8.7e-26;
  Matches  238;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         65 TTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGA 124
              ||  ||  |  |  || ||| ||||||||||||| | ||||| ||||||||||||||||
Db         66 TTCTTAAGATGTACTTGGCCGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 125

Qy        125 GACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCA 184
              | ||||||  |||||||||| ||||  |||||||| ||||||||||||||||||||||||
Db        126 GGCCGAGTCAGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGCA 185

Qy        185 AAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGT 244
              ||||  |   || || ||||| |||||||||||||| ||  || |||||  ||||| || |
Db        186 AAACCCCTCCTGTACTAAAAGTACAAAAATTAGCCTGGCTTGCTGGCAGGTGCCTGTAAT 245

Qy        245 CCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCA 304
              |||||||| |||||||||  ||||| |||||| |||||||| |||||||||||||| ||
Db        246 CCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCCCTTGAACCCGGGAGGTGGAGGTTACA 305

Qy        305 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGG 349
              |||||||||||||||||||||||||||||||||||||||||||   | ||| |
Db        306 GTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGGCGACAAGAG 350
```

```
RESULT 5
AG119835
LOCUS       AG119835                683 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-128G10.F, genomic survey sequence.
ACCESSION   AG119835
VERSION     AG119835.1  GI:16740354
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 683)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-128G10.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
```

```
  Query Match            42.0%;  Score 209.6;  DB 10;  Length 683;
```

```
  Best Local Similarity   78.4%;  Pred. No. 9.6e-26;
  Matches  251;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy         61 TCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTT 120
              | |||||| || |  | |  ||| ||  |||||||||   || || ||||||||||||
Db        183 TGAGTTAAAAGCATGTAAAAGGCCGGGCACGGTGGCTCAGACCTATAATCCCAGCACTTT 242

Qy        121 GGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACAT 180
              | ||  | |||  |||||||||| ||||| |||||||| ||||||||||||||||||||
Db        243 GAGAAGCTGAGGCGGGTGGATCACCTGAGATCAGGAGTTCGAGACCAGCCTGGCCAACAT 302

Qy        181 GGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTA 240
              ||   ||| |  |||||| |||||||||||||||||||||||| |||||| ||   ||||
Db        303 GGTGTAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGACAGGCGCCTG 362

Qy        241 TAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGT 300
              || |||||||||| || ||||||  |||||||||||||||||||||||||| |||||
Db        363 TAATCCCAGCTACTCACGAGGCTAAGGCAGAAGAATCACTTGAACCTGGGAGGTAGAGGT 422

Qy        301 TGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAAC 360
              ||||||||||||||||| || |||||||||||||||||   | ||| |  |    | |
Db        423 TGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCCTGGGGGACAAGAGCCAGACTTCAT 482

Qy        361 TTTACTTGCAACCATAGCTG 380
              || |   | ||  |  ||||
Db        483 TTCAAAAGAAAAAAAGGCTG 502


RESULT 6
B71494/c
LOCUS       B71494                   649 bp    DNA     linear   GSS 08-APR-1999
DEFINITION  RPCI11-8N9.TP RPCI-11 Homo sapiens genomic clone RPCI-11-8N9,
            genomic survey sequence.
ACCESSION   B71494
VERSION     B71494.1  GI:2710715
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-8N9.TV
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7503008"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-8N9"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             41.9%;  Score 209.4;  DB 9;  Length 649;
  Best Local Similarity   86.5%;  Pred. No. 1.1e-25;
  Matches  231;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy         82 GCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGAT 141
              ||| |||| ||||||| |   ||||| |||||||||||||||| | |||   ||  |||
Db        507 GCCGGGTGTGGTGGCTTACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGAT 448

Qy        142 CACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATA 201
              ||||||||  |||||||| |||||||||||||||||||||||||| |||  || ||||| ||
Db        447 CACTTGAGGTCAGGAGTTGGAGACCAGCCTGGCCAACATGGCCAAATCCTGTCTCTACTA 388

Qy        202 AAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGG 261
              |||||||||||||||||||| |||||||| | |||||||| |||||||||  ||||||
Db        387 AAAATACAAAAATTAGCCAGGAGTGGTGGCGCATGCCTATAATCCCAGCTACTTAGGAGG 328

Qy        262 CTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCAC 321
              ||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        327 CTGAGGCAGGAGAATTACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCAC 268

Qy        322 CACTGCACTCCAGCCTGGAAAAAAAGG 348
              |||| |||||||||||||| | | || |
Db        267 CACTACACTCCAGCCTGGGAGACAAAG 241


RESULT 7
AQ586662
LOCUS       AQ586662                 599 bp    DNA     linear   GSS 07-JUN-1999
DEFINITION  RPCI-11-448E3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-448E3,
            genomic survey sequence.
ACCESSION   AQ586662
VERSION     AQ586662.1  GI:5007701
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7671746"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-448E3"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
```

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
ORIGIN

Query Match 41.9%; Score 209.2; DB 9; Length 599;
Best Local Similarity 79.7%; Pred. No. 1.2e-25;
Matches 247; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy         67 AATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGA 126
              | ||  |  |||   ||  ||||| |||||||||| ||||| |||||||||||| ||||
Db         61 ATTAACATGTCACAGGCTGGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTAGGAGG 120

Qy        127 CCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAA 186
              ||||| || |||||||||||||||  ||||||||  ||||||||||||||||||||| |||
Db        121 CCGAGGTGAGTGGATCACTTGAGGCCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAA 180

Qy        187 ACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCC 246
              || |  ||| || |||||||| |||||||||||||||||||||||| ||| ||||| || |||
Db        181 ACCCCGTCTTTACTAAAAATTCAAAAATTAGCCAGGTGTGGTGCCACATGCCTGTAATCC 240

Qy        247 CAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGT 306
              ||||||  ||| ||||||| ||||| |||||| ||| ||||||||||||||||||||||
Db        241 CAGCTATTCAGCAGGCTGAGGCAGGAGAATCGCTTAAACCTGGGAGGTGGAGGTTGCAGT 300

Qy        307 GAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAATTAATAACTTTACT 366
              ||| |||||||||||||||||||||||||||||    || |  |  |        ||| ||
Db        301 GAGGCAAGATTGCACCACTGCACTCCAGCCTGAGCAATAGAGCAAGACTCCGTCTTCACA 360

Qy        367 TGCAACCATA 376
                ||  || |
Db        361 CACACACAAA 370
```

RESULT 8
BM679026/c
LOCUS BM679026 799 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-EO0-ahx-j-02-0-UI.s1 UI-E-EO0 Homo sapiens cDNA clone
UI-E-EO0-ahx-j-02-0-UI 3', mRNA sequence.
ACCESSION BM679026
VERSION BM679026.1 GI:18988922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 799)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 20-305, >ALU (matched compliment)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .799
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EO0-ahx-j-02-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EO0"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EO0 is a cDNA library containing the following
tissue(s): fetal eye. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGCGTATACC. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG_TISSUE=human fetal eye
TAG_LIB=UI-E-EO0
TAG_SEQ=CGCGTATACC"
ORIGIN

Query Match 41.8%; Score 208.8; DB 3; Length 799;
Best Local Similarity 77.8%; Pred. No. 1.2e-25;
Matches 252; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy         31 AGCATATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGC 90
              | ||| || ||  | ||| |||              |||| |  |  |   | ||| |||||
Db        355 AACATTTGAGAAGTCGGAAGGAAAGTATTACATTTTAAGATCAGTTTGCTGGCCGGGTGC 296

Qy         91 GGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGA 150
              ||||||||| | ||||| |||||||||||||||| ||||| | ||  ||||||||||
Db        295 GGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTAGGCTGATCACTTGAGG 236

Qy        151 GCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAA 210
              ||||||  |  |||||||||||||||||| | ||||| | ||||||| ||||||||||
Db        235 TCAGGAGCTCCAGACCAGCCTGGCCAACATAGTAAAACCCCATCTCTACTAAAAATACAA 176

Qy        211 AAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAG 270
              ||||||||||||  ||||||  || ||||  |||||||||||||  ||||||||||||||
Db        175 AAATTAGCCAGGCATGGTGGTACATGCCGGTAGTCCCAGCTACTCAGGAGGCTGGGGCAT 116

Qy        271 AAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACT 330
              ||||||  |||||||| |||||| | |||||||||||||||||||| |    |||||||||
Db        115 GAGAATCGCTTGAACCCGGGAGGCGCAGGTTGCAGTGAGCCAAGATCGTGGCACTGCACT 56

Qy        331 CCAGCCTGGAAAAAAAGGGTAATT 354
              ||||||||  | |||| |  | |
Db         55 GCAGCCTGGCGACAAAGTGAGACT 32
```

RESULT 9
AQ037381
LOCUS AQ037381 455 bp DNA linear GSS 11-JUL-1998
DEFINITION CIT-HSP-2335L20.TR CIT-HSP Homo sapiens genomic clone 2335L20,
genomic survey sequence.
ACCESSION AQ037381
VERSION AQ037381.1 GI:3303213
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map

```
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2335L20.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .455
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="2335L20"
                      /sex="Male"
                      /cell_type="Sperm"
                      /clone_lib="CIT-HSP"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
ORIGIN

  Query Match             41.7%;  Score 208.4;  DB 9;  Length 455;
  Best Local Similarity   83.7%;  Pred. No. 1.7e-25;
  Matches  236;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy         68 ATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGAC 127
              | | | || |||  |||||| || |||||||||| ||||| |||||| |||||||||| |
Db        124 AGACTCACACATAGGCCAGGCGCAGTGGCTCATGCCTGTAATCCCAGTACTTTGGGAGGC 183

Qy        128 CGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAA 187
              ||||  ||||||||||||||||  |||||||| ||||||||||||||||||||||  ||
Db        184 CGAGGCGGGTGGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGGGAAG 243

Qy        188 CACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCC 247
              | || |||||| |||||||||||||||||||||||| ||||||||    |||| || ||||
Db        244 CCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCGGGCGCCTGTAATCCC 303

Qy        248 AGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTG 307
              |||||| |||||||||| ||||| |||||| |||||||||||||| ||||||||||||
Db        304 AGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAGGTTGCAGTG 363

Qy        308 AGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGG 349
              |||| |||| |  ||||| |||||||||||||   | ||| |
Db        364 AGCCGAGATCGTGCCACTACACTCCAGCCTGGGCTACAAGAG 405


RESULT 10
F35684
LOCUS       F35684                   374 bp    mRNA    linear   EST 13-MAY-1999
DEFINITION  HSPD32531 HM3 Homo sapiens cDNA clone sH5-000005-0/D10, mRNA
            sequence.
ACCESSION   F35684
VERSION     F35684.1  GI:4821310
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A.,
            Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.
  TITLE     Identification of 4370 expressed sequence tags from a
            3'-end-specific cDNA library of human skeletal muscle by DNA
            sequencing and filter hybridization
  JOURNAL   Genome Res. 6 (1), 35-42 (1996)
   PUBMED   8681137
COMMENT     Contact: Valle G.
            CRIBI Biotechnology Centre
            University of Padua
            Via Trieste 75, 35121 Padua, Italy
            ABI Chromatograms and other information are available on WWW at
            http://grup.bio.unipd.it.
FEATURES              Location/Qualifiers
     source           1. .374
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="sH5-000005-0/D10"
                      /sex="female"
                      /tissue_type="pectoral muscle (after mastectomy)"
                      /clone_lib="HM3"
                      /note="Vector: pcDNAII (Invitrogen); Site_1: BstXI;
                      Site_2: NotI; The library was constructed by G.
                      Lanfranchi. This library is not subtracted nor normalized.
                      The first strand cDNA was primed with a biotinylated
                      oligo-dT-NotI primer
                      (5'-biotin-AACCCGGCTCGAGCGGCCGCTTTTTTTTTTTTTTTT-3'). The
                      ds cDNA was sonicated and size-selected in the range
                      350-550  bp. The 3' specific fragments were selected by
                      streptavidin coated magnetic beads, ligated to
                      non-palindromic BstXI adapters, NotI digested and
                      directionally cloned into BstXI-NotI cut pcDNAII vector."
ORIGIN

  Query Match             41.4%;  Score 206.8;  DB 8;  Length 374;
  Best Local Similarity   78.7%;  Pred. No. 3.4e-25;
  Matches  247;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         36 ATGGGACCTGGGATGGAGTTAGCGCTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGG 95
              ||  || | | ||  |||   |  |   ||||    ||   || |||||||||| |||||
Db         51 ATTAGAGCAGAGACTGAGGATGAACAATTTTAAAGTGAAAGTATGAGCCAGGTGTGGTGG 110

Qy         96 CTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGG 155
              |||| | ||||| |||||||| |||||||| | |||  |||||||||| ||||  ||||
Db        111 CTCACGCCTGTAATCCCAGCAGTTTGGGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGG 170

Qy        156 AGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAAAAATT 215
              ||||  ||||||||||||||||||||  |||| | ||||||| ||||||||||||||||
Db        171 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 230

Qy        216 AGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAA 275
              ||||||| |||||||||| ||| | || |||||||||| | |||||||| ||||| ||||
Db        231 AGCCAGGCGTGGTGGCACGTGCTTGTAATCCCAGCTACTCTGGAGGCTGAGGCAGGAGAA 290

Qy        276 TCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGC 335
              |  ||||||||  |||||  |||||| |||||||| |||| |||||| ||||||| |||
Db        291 TTGCTTGAACCCAGGAGGCAGAGGTTACAGTGAGCCGAGATGGCACCATTGCACTCTAGC 350

Qy        336 CTGGAAAAAAAAGGG 349
              ||||  || ||| |
Db        351 CTGGGCAACAAGAG 364


RESULT 11
BX470494/c
LOCUS       BX470494                 573 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686J21109_s1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686J21109 3', mRNA sequence.
ACCESSION   BX470494
VERSION     BX470494.1  GI:31664821
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
```

```
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 3' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            r1 sequence also available.
            This clone (DKFZp686J21109) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES            Location/Qualifiers
     source         1. .573
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="DKFZp686J21109"
                    /dev_stage="adult"
                    /lab_host="DH10B"
                    /clone_lib="686 (synonym: hlcc3)"
                    /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                    cDNA-collection"
ORIGIN

  Query Match             41.4%;  Score 206.6;  DB 5;  Length 573;
  Best Local Similarity   84.1%;  Pred. No. 3.2e-25;
  Matches  233;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy         76 TCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGG 135
              | ||| ||||||  |||||||||| | ||||| |||||||||||||||||| ||||| |||
Db        303 TAATTGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGG 244

Qy        136 GTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCT 195
              |||||||| ||||| | |||||||||||||||||| ||||||||| ||||   | ||||
Db        243 ATGGATCACCTGAGACCGGGAGTTTGAGACCAGCCTGACCAACATGGAAAAAACCCATCT 184

Qy        196 CTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACAC 255
              ||| |||||||||||||||| ||| ||  |||||||||| ||||| || ||||||||||  |
Db        183 CTACTAAAAATACAAAAATTGGCCGGGCATGGTGGCACATGCCTGTAATCCCAGCTACTC 124

Qy        256 AGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGA 315
              |||||||||| ||||| |||||   | |||  |||||||||||||||||| ||||||||||
Db        123 AGGAGGCTGAGGCAGGAGAATTGCCTGACTCTGGGAGGTGGAGGTTGCTGTGAGCCAAGA 64

Qy        316 TTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
              | |||||| ||||||| |||||||  || ||  | ||
Db         63 TCGCACCATTGCACTCTAGCCTGGGCAATAACAGCAA 27


RESULT 12
BM790237/c
LOCUS       BM790237                600 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0070027 S22SNU16 Homo sapiens cDNA clone S22SNU16-4-C11 5',
            mRNA sequence.
ACCESSION   BM790237
VERSION     BM790237.1  GI:19138469
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 4  row: C  column: 11
            High quality sequence stop: 600.
FEATURES            Location/Qualifiers
     source         1. .600
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="S22SNU16-4-C11"
                    /sex="F"
                    /tissue_type="Ascites"
                    /cell_type="Lymphoblast-like"
                    /cell_line="SNU-16"
                    /lab_host="DH10B"
                    /clone_lib="S22SNU16"
                    /note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
                    Site_2: NotI; The S22SNU16 library was contributed by the
                    Soares laboratory and it was constructed as described by
                    Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
                    Research 6(9): 791-806. RNA was prepared from harvested
                    cells of SNU-16 culture. SNU-16 cell was obtained from
                    Korean Cell Line Bank (KCLB). SNU-16 was established from
                    ascitic fluids of Korean patients by Park J.G. et al.
                    (1990), Cancer Res 50: 2773-2780."
ORIGIN

  Query Match             41.4%;  Score 206.6;  DB 3;  Length 600;
  Best Local Similarity   87.0%;  Pred. No. 3.2e-25;
  Matches  227;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy         79 TTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTG 138
              | ||||||| || |||||||| | ||||| |||||||||||||||||| ||||| ||||||
Db        494 TCAGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTG 435

Qy        139 GATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTA 198
              |||||||||||  |||||||||||||||||||||||||||||||||  |||| | |||||||
Db        434 GATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTA 375

Qy        199 ATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGG 258
              |||||  |||||||||||||||||  ||||||||  |||||  ||  |||||||||   ||
Db        374 CTAAAACTACAAAAATTAGCCAGGCATGGTGGCATGTGCCTCTAACCCCAGCTACTTGGG 315

Qy        259 AGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTG 318
              |||||| ||||| | |||||||||||||  |||||||||||||||||||||||||||||||
Db        314 AGGCTGAGGCAGGAAAATCACTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCCAAGATTG 255

Qy        319 CACCACTGCACTCCAGCCTGG 339
              |||||| ||||||||||||
Db        254 TGCCACTGTACTCCAGCCTGG 234


RESULT 13
BM788692/c
LOCUS       BM788692                657 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0067875 S10SNU1 Homo sapiens cDNA clone S10SNU1-20-H06 5',
            mRNA sequence.
ACCESSION   BM788692
VERSION     BM788692.1  GI:19136924
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
```

```
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 20  row: H  column: 06
            High quality sequence stop: 657.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S10SNU1-20-H06"
                     /sex="M"
                     /tissue_type="Stomach"
                     /cell_type="Lymphoblast-like"
                     /cell_line="SNU-1"
                     /lab_host="Top10F'"
                     /clone_lib="S10SNU1"
                     /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
                     Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including SfiI
                     site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized with Superscript II using SfiI
                     oligo-dT primer. After first strand synthesis, RNA was
                     degraded by NaOH treatment and cDNA was amplified by PCR
                     reaction. The PCR products were digested with SfiI and
                     cloned into DraIII- digested pME18S-FL3 vector. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
ORIGIN

  Query Match             41.4%;  Score 206.6;  DB 3;  Length 657;
  Best Local Similarity   87.0%;  Pred. No. 3.1e-25;
  Matches  227;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          79 TTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTG 138
               | ||||||| || |||||||| | ||||| |||||||||||||||||| ||||| ||||||
Db         425 TCAGCCAGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTG 366

Qy         139 GATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTA 198
               |||||||||||  ||||||||||||||||||||||||||||||||  |||| | |||||||
Db         365 GATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTA 306

Qy         199 ATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGG 258
               |||||  |||||||||||||||||  ||||||||  |||||  ||  |||||||||    ||
Db         305 CTAAAACTACAAAAATTAGCCAGGCATGGTGGCATGTGCCTCTAACCCCAGCTACTTGGG 246

Qy         259 AGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTG 318
               ||||||  |||||  | ||||||||||||  |||||||||||||||||||||||||||||||
Db         245 AGGCTGAGGCAGGAAAATCACTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCCAAGATTG 186

Qy         319 CACCACTGCACTCCAGCCTGG 339
                 |||||| ||||||||||||
Db         185 TGCCACTGTACTCCAGCCTGG 165


RESULT 14
CR960649
LOCUS       CR960649                795 bp    DNA     linear   GSS 06-JUN-2005
DEFINITION  Homo sapiens BAC end sequence of RZPDB737B062021D from genomic
            library (orig. Pieter J. de Jong library RPCI-11), genomic survey
            sequence.
ACCESSION   CR960649
VERSION     CR960649.1  GI:66991211
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   Schwarz,F., Neubert,P., Schneider,D., Peters,M. and Korn,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2005) RZPD Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH, Im Neuenheimer Feld 515, D-69120 Heidelberg,
            Germany
COMMENT     RZPD; RZPDB737B062021D; RP11-56D10;
            derived from Pieter J. de Jong library RPCI-11;
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPDB737B062021D
            RZPDLIB; (Human Genomic Set - RZPD 1.0) RZPD LIB No.737
            http://www.rzpd.de/cgi-bin/products/set.cgi?libNo=737
            http://www.rzpd.de/products/genomicset/
            Contact: Inge Arlart
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 100
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available from RZPD;
            Contact RZPD (customer.service@rzpd.de) for further information.
            Clone distribution: http://www.rzpd.de/products/genomicset/
            Seq-primer: T7 (TAA-TAC-GAC-TCA-CTA-TAG-GG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RZPDB737B062021D"
                     /sex="Male"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; RPCI-11 Human Male BAC Library"
ORIGIN

  Query Match             41.4%;  Score 206.6;  DB 11;  Length 795;
  Best Local Similarity   82.8%;  Pred. No. 2.9e-25;
  Matches  236;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          68 ATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGAC 127
               |||  || |     |||||| |||||||||||||| || || |||||||||||||||| |
Db         115 ATAAAAAATTTAGGGCCAGGCGCGGTGGCTCATGCCTATAATCCCAGCACTTTGGGAGGC 174

Qy         128 CGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAA 187
               | ||   |||||||||| ||||  ||||||||  |||| ||||||||||||||||||||||
Db         175 CAAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACGAGCCTGGCCAACATGGCAAAA 234

Qy         188 CACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCC 247
               | | |||||||  ||||||||||||||||||||||  ||||||     |||||||| ||||
Db         235 CCCCATCTCTACCAAAAATACAAAAATTAGCCAGGCATGGTGGTGGGTGCCTATAATCCC 294

Qy         248 AGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTG 307
               ||||||  | ||||||||  |||| ||||||||||||||| ||||||||||||||||||||
Db         295 AGCTACTCGGGAGGCTGAAGCAGGAGAATCACTTGAACCCGGGAGGTGGAGGTTGCAGTG 354

Qy         308 AGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGTAA 352
               |||  ||||| |  ||| |||||||||||||| || ||| || |
Db         355 AGCTAAGATCGTGCCATTGCACTCCAGCCTGGACAACAAGAGTGA 399


RESULT 15
BE169870/c
LOCUS       BE169870                514 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  CM3-HT0528-010200-086-f04 HT0528 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE169870
VERSION     BE169870.1  GI:8632591
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-HT0528-010
            200-086-f04&t3=2000-02-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0528"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

  Query Match             41.3%;  Score 206.4;  DB 2;  Length 514;
  Best Local Similarity   83.9%;  Pred. No. 3.6e-25;
  Matches  245;  Conservative    0;  Mismatches   46;  Indels    1;  Gaps    1;

Qy          59 GCTCAGTTAATAGTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACT 118
               |||  ||||| ||| |       |||||| |||||||| ||| ||||| ||||||||||
Db         502 GCTATGTTAAAAGTGAAAACAAGGCCAGGCGCGGTGGCT-ATGCCTGTAATCCCAGCACT 444

Qy         119 TTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAAC 178
               ||||||| |||||  |||  || ||||||||  ||||||||||||||||||| ||||||||
Db         443 TTGGGAGGCCGAGGCGGGCAGACCACTTGAGGACAGGAGTTTGAGACCAGCTTGGCCAAC 384

Qy         179 ATGGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCC 238
                |||||||| || |||||| | |||||||||||||||| |||||||||||  | ||||
Db         383 GTGGCAAAACCCTGTCTCTACTGAAAATACAAAAATTAGCTAGGTGTGGTGGTGCATGCC 324

Qy         239 TATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAG 298
               | || |||||||||| |||||||||| ||||  ||||||||||||||  ||||||||||
Db         323 TGTAATCCCAGCTACTCAGGAGGCTGAGGCATGAGAATCACTTGAACCCAGGAGGTGGAG 264

Qy         299 GTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAGGGT 350
               |||||||||||||||||||||||||||||||||||||||||||||||||  || |  | |
Db         263 GTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGGGCAACAGAGCT 212


Search completed: February  9, 2006, 18:10:10
Job time : 2363.93 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 2697.09 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 501
Sequence: 1 acgggctcactctgtcgcc.....tttaataaagacggggtttc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		501	100.0	8368	8	AY438977	AY438977 Homo sapi
2		501	100.0	13612	6	AX377239	AX377239 Sequence
3	c	501	100.0	122302	8	AC003982	AC003982 Homo sapi
4	c	501	100.0	220384	14	AC078926	AC078926 Homo sapi
5		429.2	85.7	3375	8	HUMPLA2A1	M22970 Human pancr
6	c	248.8	49.7	112069	8	AL356103	AL356103 Human DNA
7		239.8	47.9	153881	14	AC011793	AC011793 Homo sapi
8	c	239.8	47.9	181655	14	AC024488	AC024488 Homo sapi
9		238.6	47.6	193352	14	AC116046	AC116046 Papio ham
10		237	47.3	123253	8	AC004971	AC004971 Homo sapi
11	c	236.8	47.3	158905	8	AC005666	AC005666 Homo sapi
12	c	236.8	47.3	170892	14	AC068438	AC068438 Homo sapi
13	c	235.2	46.9	174902	8	AC099811	AC099811 Homo sapi
14		234.8	46.9	107529	8	AL357936	AL357936 Human DNA
15		234.2	46.7	150151	8	AL356115	AL356115 Human DNA
16	c	234.2	46.7	211001	8	AC018695	AC018695 Homo sapi
17	c	233.8	46.7	203134	8	AC021077	AC021077 Homo sapi
18		233.6	46.6	163331	8	AC092802	AC092802 Homo sapi
19		233.6	46.6	171641	8	AC018512	AC018512 Homo sapi
20	c	233.6	46.6	172421	14	AC013279	AC013279 Homo sapi
21		233.6	46.6	196603	8	AC087721	AC087721 Homo sapi
22		233.4	46.6	62613	14	AC025789	AC025789 Homo sapi
23	c	233.4	46.6	125625	8	DQ015919	DQ015919 Homo sapi
24	c	233.4	46.6	149605	8	AC005522	AC005522 Homo sapi
25		233.4	46.6	188117	8	AC022613	AC022613 Homo sapi
26	c	233.4	46.6	201215	8	AC019084	AC019084 Homo sapi
27		232.4	46.4	80586	8	AC012574	AC012574 Homo sapi
28	c	231.8	46.3	135996	14	AL356007	AL356007 Homo sapi
29		231.6	46.2	124645	8	HUAC004638	AC004638 Homo sapi
30		231.6	46.2	124998	8	AC026397	AC026397 Homo sapi
31	c	231.6	46.2	125099	8	AC006251	AC006251 Homo sapi
32		231.6	46.2	128877	14	AC026399	AC026399 Homo sapi
33		231.6	46.2	170691	8	AC044802	AC044802 Homo sapi
34	c	231.4	46.2	171713	14	AL356152	AL356152 Homo sapi
35	c	230.8	46.1	164125	8	AC018693	AC018693 Homo sapi
36	c	230.6	46.0	179390	14	AC115532	AC115532 Papio anu
37	c	230.4	46.0	92518	8	BX088650	BX088650 Human DNA
38	c	230.2	45.9	193766	8	AC093420	AC093420 Homo sapi
39		230.2	45.9	200881	14	AC116566	AC116566 Homo sapi
40		230	45.9	206102	8	AC027131	AC027131 Homo sapi
41		229.8	45.9	184762	14	AC023500	AC023500 Homo sapi
42	c	229.6	45.8	163183	8	AC016574	AC016574 Homo sapi
43	c	229.6	45.8	187691	8	AL139082	AL139082 Human DNA
44	c	229.2	45.7	143506	8	AC092833	AC092833 Homo sapi
45	c	229	45.7	167309	14	AC157857	AC157857 Papio anu

ALIGNMENTS

RESULT 1
AY438977
LOCUS AY438977 8368 bp DNA linear PRI 29-OCT-2003
DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
ACCESSION AY438977
VERSION AY438977.1 GI:37953284
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8368)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).
FEATURES Location/Qualifiers
source 1. .8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101. .257
/rpt_family="MIR"
/rpt_type=dispersed
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326. .538
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 354. .496

```
                /note="Region not scanned for variation"
gene            875. .6548
                /gene="PLA2G1B"
mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                /gene="PLA2G1B"
                /product="phospholipase A2, group IB (pancreas)"
CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                /gene="PLA2G1B"
                /codon_start=1
                /product="phospholipase A2, group IB (pancreas)"
                /protein_id="AAR05441.1"
                /db_xref="GI:37953285"
                /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
repeat_region   1174. .1307
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1286
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="g"
repeat_region   1319. .1588
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1343
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="a"
variation       1593
                /gene="PLA2G1B"
                /frequency="0.02"
                /replace="t"
repeat_region   1605. .1669
                /rpt_family="L2"
                /rpt_type=dispersed
misc_feature    1638. .2188
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
repeat_region   1670. .1975
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   1976. .2038
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2086. .2165
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2194. .2319
                /rpt_family="Alu"
                /rpt_type=dispersed
misc_feature    2319. .2453
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
variation       3027
                /gene="PLA2G1B"
                /frequency="0.25"
                /replace="t"
repeat_region   3165. .3431
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       3262
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3321
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
variation       3396
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="g"
variation       3630
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="c"
variation       3702
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="a"
variation       3768
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3818
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
repeat_region   3846. .3964
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   3977. .4280
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   4281. .4346
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   4347. .4634
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       4386
                /gene="PLA2G1B"
                /frequency="0.08"
                /replace="t"
repeat_region   4635. .4723
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5013
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="a"
repeat_region   5103. .5404
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5133
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5543
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5555
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       5556
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="g"
repeat_region   5577. .5629
                /rpt_family="L1"
                /rpt_type=dispersed
variation       5580
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
variation       5675
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
repeat_region   5779. .6083
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5929
```

```
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             100.0%;  Score 501;  DB 8;  Length 8368;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-131;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1374 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 1433

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1434 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 1493

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1494 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1553

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1554 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1613

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1614 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 1673

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1674 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 1733

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1734 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 1793

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 1853

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       1854 TTTTAATAAAGACGGGGTTTC 1874


RESULT 2
AX377239
LOCUS       AX377239               13612 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212562.
ACCESSION   AX377239
VERSION     AX377239.1  GI:19573528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 1 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .13612
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     variation       3845
                     /note="PS1: polymorphic base G or A"
     variation       3968
                     /note="PS2: polymorphic base A or G"
     variation       6060
                     /note="PS2: polymorphic base A or G"
     variation       6844
                     /note="PS4: polymorphic base G or A"
     variation       9531
                     /note="PS5: polymorphic base G or A"
ORIGIN

  Query Match             100.0%;  Score 501;  DB 6;  Length 13612;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-131;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4516 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 4575

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4576 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 4635

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4636 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 4695

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4696 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 4755

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4756 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 4815

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4816 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 4875

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4876 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 4935

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4936 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 4995

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       4996 TTTTAATAAAGACGGGGTTTC 5016


RESULT 3
AC003982/c
LOCUS       AC003982               122302 bp    DNA     linear   PRI 18-MAR-1999
DEFINITION  Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION   AC003982
VERSION     AC003982.1  GI:2769695
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122302)
  AUTHORS   Sulston,J.E. and Waterston,R.
```

```
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 122302)
  AUTHORS   Bradshaw,H., Wu,X. and Ozersky,P.
  TITLE     The sequence of Homo sapiens PAC clone 166H1
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was originally isolated in the laboratory of Professor
            Graeme Bell, Howard Hughes Medical Institute and Departments of
            Biochemistry and Molecular Biology, and Medicine, The University of
            Chicago, Chicago, IL, USA.  The clone was provided by the
            laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
            Genetics, Oxford, UK.  Some contig information was also obtained
            from Yamagata et al., Nature 384:455-8 (1996).

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is 278C19; the clone sequenced to
            the right is 15E1.  Actual start of this clone is at base position
            1 of 166H1; actual end is at 122302 of 166H1.
FEATURES              Location/Qualifiers
     source           1. .122302
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="12"
                      /map="12q"
                      /clone="166H1"
                      /clone_lib="RPCI-1"
     repeat_region    1. .238
                      /rpt_family="Alu"
     repeat_region    239. .265
                      /rpt_family="AT_rich"
     repeat_region    268. .565
                      /rpt_family="Alu"
     repeat_region    1001. .1309
                      /rpt_family="Alu"
     repeat_region    1440. .1745
                      /rpt_family="Alu"
     repeat_region    1909. .1940
                      /rpt_family="MER1_type"
     repeat_region    1955. .1983
                      /rpt_family="AT_rich"
     repeat_region    2063. .2362
                      /rpt_family="Alu"
     repeat_region    2516. .2650
                      /rpt_family="Alu"
     repeat_region    2674. .2730
                      /rpt_family="L1"
     repeat_region    2735. .2788
                      /rpt_family="(GAAAA)n"
     repeat_region    2790. .3077
                      /rpt_family="Alu"
     repeat_region    3081. .3139
                      /rpt_family="L1"
     repeat_region    3716. .4020
                      /rpt_family="Alu"
     repeat_region    4196. .4492
                      /rpt_family="Alu"
     gene             4560. .14034
                      /gene="WUGSC:H_166H1.1"
     CDS              join(4560. .5056,13418. .13748,13882. .14034)
                      /gene="WUGSC:H_166H1.1"
                      /note="unknown function; 60% similar to Z50177
                      (PID:g927403) (PID:g927402); H_166H1.1"
                      /codon_start=1
                      /evidence=not_experimental
                      /protein_id="AAB95634.1"
                      /db_xref="GI:2769696"
                      /translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
                      KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
                      QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
                      HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
                      LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
                      VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
     repeat_region    5366. .5655
                      /rpt_family="Alu"
     repeat_region    5672. .5798
                      /rpt_family="Alu"
     repeat_region    5802. .5834
                      /rpt_family="(TA)n"
     repeat_region    5834. .5857
                      /rpt_family="POLY_A"
     repeat_region    5870. .6184
                      /rpt_family="Alu"
     repeat_region    6224. .6268
                      /rpt_family="(CA)n"
     repeat_region    6269. .6558
                      /rpt_family="Alu"
     repeat_region    6729. .7029
                      /rpt_family="Alu"
     repeat_region    7051. .7172
                      /rpt_family="Alu"
     repeat_region    7192. .7500
                      /rpt_family="Alu"
     repeat_region    7518. .7820
                      /rpt_family="Alu"
     repeat_region    8207. .8500
                      /rpt_family="Alu"
     repeat_region    8576. .8878
                      /rpt_family="Alu"
     repeat_region    9023. .9154
                      /rpt_family="Alu"
     repeat_region    9165. .9470
                      /rpt_family="Alu"
```

```
     repeat_region   9887. .10188
                     /rpt_family="Alu"
     repeat_region   10221. .10342
                     /rpt_family="L1"
     repeat_region   10408. .10707
                     /rpt_family="Alu"
     repeat_region   10713. .11014
                     /rpt_family="Alu"
     repeat_region   11020. .11110
                     /rpt_family="Alu"
     repeat_region   11146. .11386
                     /rpt_family="Alu"
     repeat_region   11208. .11386
                     /rpt_family="Alu"
     repeat_region   11391. .11515
                     /rpt_family="Alu"
     repeat_region   11568. .11705
                     /rpt_family="Alu"
     repeat_region   11737. .12032
                     /rpt_family="Alu"
     repeat_region   12034. .12066
                     /rpt_family="7SLRNA"
     repeat_region   12075. .12372
                     /rpt_family="Alu"
     repeat_region   12448. .12753
                     /rpt_family="Alu"
     repeat_region   12763. .12874
                     /rpt_family="L2"
     repeat_region   12999. .13303
                     /rpt_family="Alu"
     misc_feature    complement(13674. .13752)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     misc_feature    complement(13840. .14195)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     repeat_region   14196. .14229
                     /rpt_family="AT_rich"
     repeat_region   14240. .14339
                     /rpt_family="U6"
     repeat_region   14349. .14646
                     /rpt_family="Alu"
     repeat_region   14652. .14940
                     /rpt_family="Alu"

  Query Match             100.0%;  Score 501;  DB 8;  Length 122302;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-131;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28288 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 28229

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28228 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 28169

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28168 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 28109

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28108 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 28049

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28048 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 27989

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      27988 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 27929

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      27928 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 27869

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      27868 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 27809

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db      27808 TTTTAATAAAGACGGGGTTTC 27788


RESULT 4
AC078926/c
LOCUS       AC078926              220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match             100.0%;  Score 501;  DB 14;  Length 220384;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-131;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73503 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 73444

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73443 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 73384

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73383 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 73324

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73323 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 73264

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73263 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 73204

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73203 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 73144

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73143 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 73084

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73083 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 73024

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db      73023 TTTTAATAAAGACGGGGTTTC 73003


RESULT 5
HUMPLA2A1
LOCUS       HUMPLA2A1               3375 bp    DNA     linear   PRI 07-JAN-1995
DEFINITION  Human pancreatic phospholipase A-2 (PLA-2) gene, exons 1 to 3.
ACCESSION   M22970 M14965
VERSION     M22970.1  GI:190008
KEYWORDS    phospholipase A2.
SEGMENT     1 of 2
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3375)
  AUTHORS   Seilhamer,J.J., Randall,T.L., Yamanaka,M. and Johnson,L.K.
  TITLE     Pancreatic phospholipase A2: isolation of the human gene and cDNAs
            from porcine pancreas and human lung
  JOURNAL   DNA 5 (6), 519-527 (1986)
```

```
   PUBMED   3028739
COMMENT     Original source text: Homo sapiens pulverized liver DNA.
FEATURES             Location/Qualifiers
     source          1. .3375
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="12q23-qter"
                     /tissue_type="pulverized liver"
     exon            <537. .570
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=1
     intron          571. .2262
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=1
     exon            2263. .2422
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=2
     intron          2423. .3222
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=2
     exon            3223. .3350
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=3
     intron          3351. .>3375
                     /gene="PLA2A"
                     /note="G00-120-715"
                     /number=3
ORIGIN      5 bp upstream of PstI site.

  Query Match             85.7%;  Score 429.2;  DB 8;  Length 3375;
  Best Local Similarity   97.4%;  Pred. No. 8.5e-111;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||| |||||||||||||||||||  |||||||||||||||||||||
Db        996 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1054

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||||| ||||||||||||||||||||||||||||||||||| ||||||||
Db       1055 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1172

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1232

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              |||||||||||||||||||||||| |||||||||||||||||||||||||     |||
Db       1233 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1292

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1293 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1352

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1412

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1413 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 1471

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       1472 TTTTTAATAAAGACGGGGTTTC 1493


RESULT 6
AL356103/c
LOCUS       AL356103              112069 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP4-713A24 on chromosome 1p22.2-31.1
            Contains STSs and GSSs, complete sequence.
ACCESSION   AL356103
VERSION     AL356103.8  GI:10178511
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 112069)
  AUTHORS   Corby,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 15, 2000 this sequence version replaced gi:9988355.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-713A24 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact:, vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .112069
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704A24713"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p22.2-31.1"
                     /clone="RP4-713A24"
                     /clone_lib="RPCI-4"
     misc_feature    1
                     /note="Clone_left_end: RP4-713A24"
     misc_feature    112069
                     /note="Clone_right_end: RP4-713A24"
ORIGIN

  Query Match             49.7%;  Score 248.8;  DB 8;  Length 112069;
  Best Local Similarity   73.4%;  Pred. No. 1.6e-59;
  Matches  367;  Conservative    0;  Mismatches  112;  Indels   21;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | |||||||||| |||||||||||||||  ||||| | || || | ||||| ||||||
Db      65537 GAGTCTCACTCTGTTGCCCAGGCTGGAGTGTAGTGGCATAATTTCGGCTCACTGCAACCT 65478
```

```
Qy        63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
             ||  |||||||||||||  ||||||| ||||||||||| | ||||||||||||
Db     65477 CCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 65423

Qy       123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                 | ||||||  ||||||| ||||||||||||||||      | ||||| ||
Db     65422 ----------ACAGGCACATGCCACCATACCCAGCTAATTTTTTG-----TATTTTTAGT 65378

Qy       183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
             |||||||||  |||||||||||| |||||| |||      || | |      |      ||
Db     65377 AGAGACAGGGTTTCACCATGTTAGCCAGGATGGTCTGAATCTCCTGACCTTGTAATCCAC 65318

Qy       243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
             ||  ||    |   |    | ||        | |     ||    |   |    |  |
Db     65317 CCGCCTTGGCTTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACGGTGCCCGGCCTGGTC 65258

Qy       303 AT-TATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTG 361
             || ||||||| || || || |||| |||||| ||| |||||||| |||||||||| |||||
Db     65257 ATGTATTCTTTTTTTTTTTTAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTG 65198

Qy       362 GTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAG 421
             | |||||||| ||||||||||||||  ||| ||||||||||||||||||| |||||||
Db     65197 GCGCGATCTCAGCTCACTGCAAGCTCTGCCTCCTGGGTTCACGCCATTCTCCTGCCTCAG 65138

Qy       422 CCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATT 481
             |||||| |||||||||||||| ||| ||||||||||||||| ||||||||||||||||
Db     65137 CCTCCCGAGTAGCTGGGACTACAGGTGCCTGCCACCACGCCCAGCTAATTTTTTTGTATT 65078

Qy       482 TTTAATAAAGACGGGGTTTC 501
             |||| || ||||||||||||
Db     65077 TTTAGTAGAGACGGGGTTTC 65058
```

RESULT 7
AC011793
LOCUS AC011793 153881 bp DNA linear HTG 27-APR-2000
DEFINITION Homo sapiens clone RP11-17K13, WORKING DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION AC011793
VERSION AC011793.3 GI:7654379
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153881)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-17K13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153881)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 27, 2000 this sequence version replaced gi:6446861.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L3652
Center clone name: 17_K_13
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148446 bases at least Q40
Consensus quality: 150611 bases at least Q30
Consensus quality: 151834 bases at least Q20
Insert size: 159000; agarose-fp
Insert size: 153281; sum-of-contigs
Quality coverage: 5.5 in Q20 bases; agarose-fp
Quality coverage: 5.7 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4883: contig of 4883 bp in length
* 4884 4983: gap of 100 bp
* 4984 10971: contig of 5988 bp in length
* 10972 11071: gap of 100 bp
* 11072 18095: contig of 7024 bp in length
* 18096 18195: gap of 100 bp
* 18196 35610: contig of 17415 bp in length
* 35611 35710: gap of 100 bp
* 35711 52825: contig of 17115 bp in length
* 52826 52925: gap of 100 bp
* 52926 92804: contig of 39879 bp in length
* 92805 92904: gap of 100 bp
* 92905 153881: contig of 60977 bp in length.

FEATURES Location/Qualifiers
source 1. .153881
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-17K13"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature 1. .4883
/note="assembly_fragment"
gap 4884. .4983
/estimated_length=100
misc_feature 4984. .10971
/note="assembly_fragment"
gap 10972. .11071
/estimated_length=100
misc_feature 11072. .18095
/note="assembly_fragment"
gap 18096. .18195
/estimated_length=100
misc_feature 18196. .35610
/note="assembly_fragment"
gap 35611. .35710
/estimated_length=100
misc_feature 35711. .52825
/note="assembly_fragment
clone_end:SP6
vector_side:left"
gap 52826. .52925
/estimated_length=100
misc_feature 52926. .92804
/note="assembly_fragment"
gap 92805. .92904
/estimated_length=100

```
     misc_feature    92905. .153881
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             47.9%;  Score 239.8;  DB 14;  Length 153881;
  Best Local Similarity   72.3%;  Pred. No. 5.8e-57;
  Matches  362;  Conservative    0;  Mismatches  117;  Indels   22;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | |||||||||| ||||||||||||||||  ||||| | || || | ||||| ||||||
Db      84323 GAGTCTCACTCTGTTGCCCAGGCTGGAGTGTAGTGGCATAATTTCGGCTCACTGCAACCT 84382

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||||||||||||  ||||||| |||||||||||| | ||||||||||||
Db      84383 CCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 84437

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                    | ||||||  |||||||| ||||||||||||||||     | ||||| ||
Db      84438 ----------ACAGGCACATGCCACCATACCCAGCTAATTTTTTG-----TATTTTTAGT 84482

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||||| |||||||||||| |||||| |||       || | |      |       ||
Db      84483 AGAGACAGGGTTTCACCATGTTAGCCAGGATGGTCTGAATCTCCTGACCTTGTAATCCAC 84542

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC--ACATTTA 300
              ||  ||    |   |    | ||      | |     ||    |  |   |  |   |
Db      84543 CCGCCTTGGCTTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACGGTGCCCGGCCTGGTC 84602

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              | |  |  |  || || || |||| |||||| ||| ||||||| |||||||||| ||||
Db      84603 ATGTATTCTTCTTTTTTTTTTAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGT 84662

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              || |||||||| |||||||||||||||  ||| ||||||||||||||||||| ||||||
Db      84663 GGCGCGATCTCAGCTCACTGCAAGCTCTGCCTCCTGGGTTCACGCCATTCTCCTGCCTCA 84722

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||| |||||||||||||| ||| ||||||||||||||| ||||||||||||||||
Db      84723 GCCTCCCGAGTAGCTGGGACTACAGGTGCCTGCCACCACGCCCAGCTAATTTTTTTGTAT 84782

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              ||||| || ||||||||||||
Db      84783 TTTTAGTAGAGACGGGGTTTC 84803


RESULT 8
AC024488/c
LOCUS       AC024488              181655 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-137A6 map 1, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC024488
VERSION     AC024488.3  GI:8954147
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 181655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-137A6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7249343.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7151
                Center clone name: 137_A_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167600 bases at least Q40
                Consensus quality: 175084 bases at least Q30
                Consensus quality: 178093 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 179755; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1308: contig of 1308 bp in length
            *     1309     1408: gap of 100 bp
            *     1409     2793: contig of 1385 bp in length
            *     2794     2893: gap of 100 bp
            *     2894     3988: contig of 1095 bp in length
            *     3989     4088: gap of 100 bp
            *     4089     5324: contig of 1236 bp in length
            *     5325     5424: gap of 100 bp
            *     5425     7027: contig of 1603 bp in length
            *     7028     7127: gap of 100 bp
            *     7128     8482: contig of 1355 bp in length
            *     8483     8582: gap of 100 bp
            *     8583    11864: contig of 3282 bp in length
            *    11865    11964: gap of 100 bp
            *    11965    15832: contig of 3868 bp in length
            *    15833    15932: gap of 100 bp
            *    15933    20128: contig of 4196 bp in length
            *    20129    20228: gap of 100 bp
            *    20229    26190: contig of 5962 bp in length
            *    26191    26290: gap of 100 bp
            *    26291    34185: contig of 7895 bp in length
            *    34186    34285: gap of 100 bp
            *    34286    42981: contig of 8696 bp in length
            *    42982    43081: gap of 100 bp
            *    43082    50160: contig of 7079 bp in length
```

```
          *    50161    50260: gap of 100 bp
          *    50261    60943: contig of 10683 bp in length
          *    60944    61043: gap of 100 bp
          *    61044    70244: contig of 9201 bp in length
          *    70245    70344: gap of 100 bp
          *    70345    81838: contig of 11494 bp in length
          *    81839    81938: gap of 100 bp
          *    81939    90735: contig of 8797 bp in length
          *    90736    90835: gap of 100 bp
          *    90836   105642: contig of 14807 bp in length
          *   105643   105742: gap of 100 bp
          *   105743   123453: contig of 17711 bp in length
          *   123454   123553: gap of 100 bp
          *   123554   181655: contig of 58102 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181655
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-137A6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1308
                     /note="assembly_fragment"
     gap             1309. .1408
                     /estimated_length=100
     misc_feature    1409. .2793
                     /note="assembly_fragment"
     gap             2794. .2893
                     /estimated_length=100
     misc_feature    2894. .3988
                     /note="assembly_fragment"
     gap             3989. .4088
                     /estimated_length=100
     misc_feature    4089. .5324
                     /note="assembly_fragment"
     gap             5325. .5424
                     /estimated_length=100
     misc_feature    5425. .7027
                     /note="assembly_fragment"
     gap             7028. .7127
                     /estimated_length=100
     misc_feature    7128. .8482
                     /note="assembly_fragment"
     gap             8483. .8582
                     /estimated_length=100
     misc_feature    8583. .11864
                     /note="assembly_fragment"
     gap             11865. .11964
                     /estimated_length=100
     misc_feature    11965. .15832
                     /note="assembly_fragment"
     gap             15833. .15932
                     /estimated_length=100
     misc_feature    15933. .20128
                     /note="assembly_fragment"
     gap             20129. .20228
                     /estimated_length=100
     misc_feature    20229. .26190
                     /note="assembly_fragment"
     gap             26191. .26290
                     /estimated_length=100
     misc_feature    26291. .34185
                     /note="assembly_fragment"
     gap             34186. .34285
                     /estimated_length=100
     misc_feature    34286. .42981
                     /note="assembly_fragment"
     gap             42982. .43081
                     /estimated_length=100
     misc_feature    43082. .50160
                     /note="assembly_fragment"
     gap             50161. .50260
                     /estimated_length=100
     misc_feature    50261. .60943
                     /note="assembly_fragment"
     gap             60944. .61043
                     /estimated_length=100
     misc_feature    61044. .70244
                     /note="assembly_fragment"
     gap             70245. .70344
                     /estimated_length=100
     misc_feature    70345. .81838
                     /note="assembly_fragment"
     gap             81839. .81938
                     /estimated_length=100
     misc_feature    81939. .90735
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             90736. .90835
                     /estimated_length=100
     misc_feature    90836. .105642
                     /note="assembly_fragment"
     gap             105643. .105742
                     /estimated_length=100
     misc_feature    105743. .123453
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     gap             123454. .123553
                     /estimated_length=100

  Query Match             47.9%;  Score 239.8;  DB 14;  Length 181655;
  Best Local Similarity   72.3%;  Pred. No. 5.9e-57;
  Matches  362;  Conservative    0;  Mismatches  117;  Indels   22;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | |||||||||| ||||||||||||||||  ||||| | || || | ||||| ||||||
Db     120547 GAGTCTCACTCTGTTGCCCAGGCTGGAGTGTAGTGGCATAATTTCGGCTCACTGCAACCT 120488

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||||||||||||  ||||||| |||||||||||| | ||||||||||||||
Db     120487 CCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 120433

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | ||||||  ||||||| ||||||||||||||||     | ||||| ||
Db     120432 ----------ACAGGCACATGCCACCATACCCAGCTAATTTTTTG-----TATTTTTAGT 120388

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||||| |||||||||||| |||||| |||      || | |     |      ||
Db     120387 AGAGACAGGGTTTCACCATGTTAGCCAGGATGGTCTGAATCTCCTGACCTTGTAATCCAC 120328

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC--ACATTTA 300
              ||  ||    |   |    | ||          |  |     ||    |   |   |    |
Db     120327 CCGCCTTGGCTTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACGGTGCCCGGCCTGGTC 120268

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              | |   |  |  || || || |||| |||||| ||| ||||||| |||||||||| ||||
Db     120267 ATGTATTCTTCTTTTTTTTTTAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGT 120208

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              || |||||||| |||||||||||||||  ||| ||||||||||||||||||| ||||||
Db     120207 GGCGCGATCTCAGCTCACTGCAAGCTCTGCCTCCTGGGTTCACGCCATTCTCCTGCCTCA 120148

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||| |||||||||||||| ||| ||||||||||||||| ||||||||||||||||
Db     120147 GCCTCCCGAGTAGCTGGGACTACAGGTGCCTGCCACCACGCCCAGCTAATTTTTTTGTAT 120088

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              ||||| || ||||||||||||
Db     120087 TTTTAGTAGAGACGGGGTTTC 120067

RESULT 9
```

```
AC116046
LOCUS       AC116046               193352 bp    DNA     linear   HTG 04-FEB-2004
DEFINITION  Papio hamadryas clone RP41-343P17, *** SEQUENCING IN PROGRESS ***,
            4 ordered pieces.
ACCESSION   AC116046
VERSION     AC116046.2  GI:42270640
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Papio hamadryas (hamadryas baboon)
  ORGANISM  Papio hamadryas
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 193352)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Papio hamadryas, clone RP41-343P17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193352)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 193352)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 4, 2004 this sequence version replaced gi:19697485.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12531
                Center clone name: 343_P_17
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    41116: contig of 41116 bp in length
            *   41117    41216: gap of 100 bp
            *   41217   127557: contig of 86341 bp in length
            *  127558   127657: gap of 100 bp
            *  127658   157700: contig of 30043 bp in length
            *  157701   157800: gap of 100 bp
            *  157801   193352: contig of 35552 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193352
                     /organism="Papio hamadryas"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9557"
                     /clone="RP41-343P17"
                     /clone_lib="RPCI-41 Male Baboon BAC"
     gap             41117. .41216
                     /estimated_length=100
     gap             127558. .127657
                     /estimated_length=100
     gap             157701. .157800
                     /estimated_length=100
ORIGIN

  Query Match           47.6%;  Score 238.6;  DB 14;  Length 193352;
  Best Local Similarity  70.7%;  Pred. No. 1.3e-56;
  Matches  353;  Conservative    0;  Mismatches  129;  Indels   17;  Gaps    2;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| |||||| |||||||||||||||| |||||  |||||||| ||||| ||| ||
Db     114678 GAGTCTCGCTCTGTAGCCCAGGCTGGAGTGCAGTGGCCGGATCTCAGCTCACTGCAAGCT 114737

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||| |||||||     ||||||  |||||||||| | ||||||||||||||
Db     114738 CCGCCTCCCGGGTTCACGCCATTCTCCGGCCTCAGCCTCCCGAGTAGCTGGGACT----- 114792

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | ||| | | |||||  |||| |||| |||||| | | ||||| || ||
Db     114793 ----------ACAGGCGCCCGCCACCTCGCCCGGCTATTTTTTTTTTTTTTTTTCTTAGT 114842

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              | |||| || ||||||| |||| |||||| |||      |   ||  |    ||
Db     114843 AAAGACGGGGTTTCACCGTGTTAGCCAGGATGG--TCTCGATCTCCTGACCTCGTGATCC 114900

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              |   |||   | ||||          | | ||      ||     ||    | |
Db     114901 GCCCATCTCGGCCTCCCAAAGTGCTGGGATTACAGGCTTGAGCCACCGCGCCCGGCCTGA 114960

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              || || || || || ||||||| |||||| ||  |||||| |||||||||| ||||||
Db     114961 CTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCGCTGTCGCCCAGGCTGGAGTGCAGTGG 115020

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              ||||||| |||||||||||||||| ||| | |||||||||||||||||| ||||||||
Db     115021 CGCGATCTTGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGC 115080

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||| |||| |||||||  ||||| ||||| ||||||||||||||||||
Db     115081 CTCCTGAGTAGCTGGTACTACAGGCGCCCACCACCGCGCCCGGCTAATTTTTTTGTATTT 115140
```

```
Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db     115141 TTAGTAGAGACGGGGTTTC 115159
```

RESULT 10
AC004971
LOCUS AC004971 123253 bp DNA linear PRI 26-APR-2003
DEFINITION Homo sapiens PAC clone RP5-1125K23 from 7, complete sequence.
ACCESSION AC004971
VERSION AC004971.3 GI:10047915
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 123253)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 123253)
AUTHORS Ali,J., Miller,K. and Strowmatt,C.
TITLE The sequence of Homo sapiens PAC clone RP5-1125K23
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 123253)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 123253)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 123253)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 123253)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 10, 2000 this sequence version replaced gi:7630775.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ1125K23
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at http://www.chori.org using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP4-607J2 the clone sequenced to the right is RP5-826E18, 200 base pair overlap. Actual end of this clone is at base position 123057 of RP5-1125K23.

FEATURES Location/Qualifiers
source 1. .123253
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP5-1125K23"
/clone_lib="RPCI-5"
gene complement(<68. .9675)
/gene="EIF3S9"
CDS complement(join(<68. .225,837. .965,1470. .1527,1833. .1952, 3695. .3887,9177. .9675))
/gene="EIF3S9"
/note="Homo sapiens eukaryotic translation initiation factor 3, subunit 9 eta, 116kDa (EIF3S9), mRNA.; H_DJ1125K23.1
Continues as H_DJ0607J02.2"
/codon_start=1
/product="unknown"
/protein_id="AAP21880.1"
/db_xref="GI:30142021"
/translation="MQDAENVAVPEAAEERAEPGQQQPAAEPPPAEGLLRPAGPGAPE
AAGTEASSEEVGIAEAGPESEVRTEPAAEAEAASGPSESPSPPAAEELPGSHAEPPVP
AQGEAPGEQARDERSDSRAQAVSEDAGGNEGRAAEAEPRALENGDADEPSFSDPEDFV
DDVSEEELLGDVLKDRPQEADGIDSVIVVDNVPQVGPDRLEKLKNVIHKIFSKFGKIT
NDFYPEEDGKTKGYIFLEYASPAHAVDAVKNADGYKLDKQHTFRVNLFTDFDKYMTIS
DEWDIPEKQPFKDLGNLRYWLEEAECRDQYSVIFESGDRTSIFWNDVKDPVSIEERAR
WTETYVRWSPKGTYLATFHQRGIALWGGEKFKQIQRFSHQGVQLIDFSPCE"
repeat_region 325. .617
/rpt_family="Alu"
repeat_region 1113. .1425
/rpt_family="Alu"
repeat_region 2194. .2331
/rpt_family="L1"
repeat_region 2338. .2644
/rpt_family="Alu"
repeat_region 2656. .2793
/rpt_family="Alu"
repeat_region 2798. .2821
/rpt_family="(CAAA)n"
repeat_region 2879. .3182
/rpt_family="Alu"
repeat_region 3317. .3484
/rpt_family="Alu"
variation 3926
/gene="EIF3S9"
/allele="T"
/allele="C"
/db_xref="dbSNP:953770"
repeat_region 4752. .5441

```
                     /rpt_family="ERV1"
     repeat_region   5526. .5656
                     /rpt_family="Alu"
     repeat_region   5769. .6067
                     /rpt_family="Alu"
     repeat_region   6154. .6190
                     /rpt_family="Alu"
     repeat_region   6220. .6500
                     /rpt_family="Alu"
     repeat_region   7547. .7687
                     /rpt_family="MIR"
     repeat_region   7779. .7961
                     /rpt_family="Alu"
     repeat_region   7965. .8266
                     /rpt_family="Alu"
     misc_feature    9058. .10603
                     /note="CpG_island (%GC=66.6, o/e=0.99, #CpGs=168)"
     repeat_region   9535. .9607
                     /rpt_family="GC_rich"
     variation       9916
                     /allele="C"
                     /allele="G"
                     /db_xref="dbSNP:2301943"
     variation       9936
                     /allele="T"
                     /allele="C"
                     /db_xref="dbSNP:2301942"
     variation       10600
                     /allele="C"
                     /allele="A"
                     /db_xref="dbSNP:2398661"
     repeat_region   10697. .10790
                     /rpt_family="Alu"
     repeat_region   10821. .11015
                     /rpt_family="L1"
     repeat_region   11016. .11148
                     /rpt_family="Alu"
     repeat_region   11149. .11182
                     /rpt_family="L1"
     repeat_region   11189. .11375
                     /rpt_family="Alu"
     repeat_region   11380. .11412
                     /rpt_family="L1"
     repeat_region   11414. .11581
                     /rpt_family="Alu"
     repeat_region   11672. .11744
                     /rpt_family="MIR"
     repeat_region   11745. .12044
                     /rpt_family="Alu"
     repeat_region   12048. .12282
                     /rpt_family="MER2_type"
     repeat_region   12283. .12383
                     /rpt_family="MaLR"
     repeat_region   12386. .12691
                     /rpt_family="Alu"
     repeat_region   12720. .13046
                     /rpt_family="Alu"
     repeat_region   13049. .13321
                     /rpt_family="Alu"
     repeat_region   13461. .13655
                     /rpt_family="Alu"

  Query Match             47.3%;  Score 237;  DB 8;  Length 123253;
  Best Local Similarity   72.1%;  Pred. No. 3.6e-56;
  Matches  360;  Conservative    0;  Mismatches  115;  Indels   24;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | |  |||| ||||| |||||||||||||||| |||| |||||||     ||||| ||||||
Db      28501 GAGTTTCACACTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTTGACTCACTGCAACCT 28560

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
                |    ||| ||||||||| ||||||| || ||||||||||||||||| ||||||
Db      28561 CTGCGTCCAGGGTTCAAACAATTCTCCTGTCTCAGCCTCTCAAGTATCTGGGA------- 28613

Qy        123 CTCTCAAGTAGCTGGCACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                |  | | | |    |||||||||||||||||||| |        |||| ||
Db      28614 --------TGACAGACGCCTGACACCATGCCCAGCTAATTTTTAG-----CAATTTTAGT 28660

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||  || ||||||||||||||||||||||||      ||   |||  |   |   |
Db      28661 AGAGATGGGGTTTCACCATGTTGGCCAGGCTGGCATTATTCTGATTTCTTTCACCACAAA 28720

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
                |  |  | ||    | |  ||| |  |      || |||    ||  |  ||| ||
Db      28721 TTAACATTCATTTT----TCTGAAATTCCATCAAGTTTTGCCTACTGTAGAATTTTTTTT 28776

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              || || || || || ||||||||||||||| ||| ||||||| |||||||||| |||||
Db      28777 TTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGC 28836

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              ||| |||||||||| |||||||| ||| |  |||||||||||||| || ||||||||
Db      28837 CGCGGTCTCGGCTCACCGCAAGCTCCGCCTCCCAGGTTCACGCCATTCGCCTGCCTCAGC 28896

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||| ||| |||||||||| ||||||||||||||||||||| ||||||||||| |||||
Db      28897 CTCCCAAGTGGCTGGGACTACAGGCGCCTGCCACCACGCCCAGCTAATTTTTTTCTATTT 28956

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db      28957 TTAGTAGAGACGGGGTTTC 28975


RESULT 11
AC005666/c
LOCUS       AC005666               158905 bp    DNA     linear   PRI 22-DEC-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.112_H_10, complete sequence.
ACCESSION   AC005666
VERSION     AC005666.1  GI:4049331
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 158905)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.112_H_10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158905)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W.,
            Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G.,
            Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
            Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczky,J.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
            Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158905)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
```

```
            Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
            Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
            McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
            Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 22, 1998 this sequence version replaced gi:4028945.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .158905
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPK.112_H_10"
                     /clone_lib="RPCI-11 human BAC library"
     repeat_region   complement(835. .1135)
                     /rpt_family="AluSq"
     repeat_region   complement(1624. .1918)
                     /rpt_family="AluSg"
     repeat_region   2302. .2467
                     /rpt_family="L1MB3"
     repeat_region   2471. .2778
                     /rpt_family="AluY"
     repeat_region   2779. .2925
                     /rpt_family="AluJo/FRAM"
     repeat_region   2926. .2948
                     /rpt_family="(CAAA)n"
     repeat_region   2955. .3255
                     /rpt_family="L1MB3"
     repeat_region   3256. .3560
                     /rpt_family="AluSc"
     repeat_region   3570. .3729
                     /rpt_family="FAM"
     repeat_region   3730. .3787
                     /rpt_family="L1MB3"
     repeat_region   complement(3964. .4114)
                     /rpt_family="MER5B"
     repeat_region   5111. .5426
                     /rpt_family="AluJb"
     repeat_region   5480. .5515
                     /rpt_family="AT_rich"
     repeat_region   5752. .5784
                     /rpt_family="AT_rich"
     repeat_region   complement(5999. .6328)
                     /rpt_family="MER58B"
     repeat_region   6435. .6518
                     /rpt_family="L1MB5"
     repeat_region   6519. .6815
                     /rpt_family="AluY"
     repeat_region   6816. .7533
                     /rpt_family="L1MB5"
     repeat_region   7535. .7694
                     /rpt_family="AluJb"
     repeat_region   7705. .8001
                     /rpt_family="AluSq"
     repeat_region   8124. .8183
                     /rpt_family="MLT2D"
     repeat_region   complement(8184. .8205)
                     /rpt_family="AluSq"
     repeat_region   complement(8206. .8229)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(8230. .8515)
                     /rpt_family="AluSq"
     repeat_region   8516. .8942
                     /rpt_family="MLT2D"
     repeat_region   complement(8957. .9079)
                     /rpt_family="MER7A"
     repeat_region   9080. .9419
                     /rpt_family="THE1C"
     repeat_region   complement(9420. .9626)
                     /rpt_family="MER7A"
     repeat_region   9702. .9748
                     /rpt_family="MLT2D"
     repeat_region   9863. .9940
                     /rpt_family="L2"
     repeat_region   10022. .10238
                     /rpt_family="L1MB3"
     repeat_region   complement(10325. .11142)
                     /rpt_family="L2"
     repeat_region   complement(11143. .11173)
                     /rpt_family="(TGAA)n"
     repeat_region   complement(11175. .11440)
                     /rpt_family="AluJb"
     repeat_region   complement(11441. .12055)
                     /rpt_family="L2"
     repeat_region   complement(12056. .12364)
                     /rpt_family="AluSq"
     repeat_region   complement(12365. .12447)
                     /rpt_family="L2"
     repeat_region   12448. .12722
                     /rpt_family="AluSc"
     repeat_region   12726. .13024
                     /rpt_family="AluJb"
     repeat_region   complement(13025. .13417)
                     /rpt_family="L2"
     repeat_region   13419. .13698
                     /rpt_family="AluSx"
     repeat_region   13699. .13726
                     /rpt_family="(TAAA)n"
     repeat_region   13783. .14076
                     /rpt_family="AluY"
     repeat_region   complement(14083. .14322)
                     /rpt_family="MIR"
     repeat_region   14484. .14510
                     /rpt_family="AT_rich"
     repeat_region   complement(15900. .16212)
                     /rpt_family="L1MA4A"
     repeat_region   16280. .16596
                     /rpt_family="MER7A"
     repeat_region   16986. .17306
                     /rpt_family="AluJo"
     repeat_region   17422. .17468
                     /rpt_family="(CAAAAA)n"
     repeat_region   17550. .17642
                     /rpt_family="HAL1"
     repeat_region   complement(17643. .17937)
                     /rpt_family="AluSp"
     repeat_region   17938. .18133
                     /rpt_family="HAL1"
     repeat_region   18764. .19102
                     /rpt_family="L1ME"
     repeat_region   19121. .19742
                     /rpt_family="L1M4"
     repeat_region   20318. .20493
                     /rpt_family="MLT1J"
     repeat_region   complement(20498. .20831)
                     /rpt_family="L1M4"
     repeat_region   complement(20888. .21059)
                     /rpt_family="AluSg"
     repeat_region   complement(21060. .21328)
                     /rpt_family="AluY"
     repeat_region   complement(21329. .21460)
                     /rpt_family="AluSg"
     repeat_region   22123. .23454
                     /rpt_family="L1MB2"
     repeat_region   23455. .23586
                     /rpt_family="FLAM_C"
```

```
     repeat_region   23587. .24009
                     /rpt_family="L1MB2"
     repeat_region   24010. .24312
                     /rpt_family="AluSq"
     repeat_region   24333. .25000
                     /rpt_family="L1M1"
     repeat_region   25001. .25298
                     /rpt_family="AluSg"
     repeat_region   25299. .26025
                     /rpt_family="L1M1"
     repeat_region   26026. .26321
                     /rpt_family="AluSx"
     repeat_region   26335. .27573

  Query Match            47.3%;  Score 236.8;  DB 8;  Length 158905;
  Best Local Similarity  73.0%;  Pred. No. 4.2e-56;
  Matches  371;  Conservative    0;  Mismatches  107;  Indels   30;  Gaps    4;

Qy          5 GGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCC 64
              | |||||||| || ||||||||||||||  ||||| |||||  | |||||| ||| |||
Db      28486 GTCTCACTCTCTCACCCAGGCTGGAGTGCAATGGCGCGATCTTGGCTCACTGCAAACTCT 28427

Qy         65 ATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCT 124
                |||||||||||||||| |||||| |||||||||||| ||||||||||||
Db      28426 GCCTCCTGGGTTCAAATGATTCTCGTGCCTCAGCCTCCCAAGTAGCTGGGA--------- 28376

Qy        125 CTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAG 184
                  | | || |||| |||||| ||||||||||||||||   ||| ||||||||||
Db      28375 ------TTACAGGTACACGCCACCAAGCCCAGCTAATTTTTT---TGTATTTTTTGGTAG 28325

Qy        185 AGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCC----------TTTTAGAT 233
              | ||||  |||||||||||||||||||  |        | |          | |
Db      28324 AAACAGAATTTCACCATGTTGGCCAGGCCAGTCTCAAACTCCTGACCTCAAGTGATCTGC 28265

Qy        234 TCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC 293
                |||   ||| |   | |  |    || |   ||      | | || |||   ||   |
Db      28264 CTGCCTTGGCCTCCCACAGTGCTAGGATTACAGGTGTGAGCTACCGTGTCCAGCCTATGC 28205

Qy        294 ACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGA 353
              | || | || || || || || || ||||||| |||||| ||| |||| || ||||||||
Db      28204 ATATCTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCACCCAGGCTGGA 28145

Qy        354 GTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCC 413
              || |||||| | |||||||||||||||||||||||||  || | ||||||||||||||||||
Db      28144 GTGCAGTGGCGGGATCTCGGCTCACTGCAAGCTCCGTCTCCCGGGTTCACGCCATTCTCC 28085

Qy        414 CGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTT 473
              |||||||||||||| |||||||||||||| ||||||| ||||| |||||| ||||| |||
Db      28084 TGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCCGCCACTACGCCCGGCTAA-TTT 28026

Qy        474 TTTGTATTTTTAATAAAGACGGGGTTTC 501
              |||||||||||| || ||||||||||||
Db      28025 TTTGTATTTTTAGTAGAGACGGGGTTTC 27998


RESULT 12
AC068438/c
LOCUS       AC068438               170892 bp    DNA     linear   HTG 24-MAY-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-110F1 map 17, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC068438
VERSION     AC068438.4  GI:21166336
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 170892)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-110F1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170892)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 170892)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 24, 2002 this sequence version replaced gi:12958069.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7824
                Center clone name: 110_F_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 3% of reads
                Sequencing vector: Plasmid; n/a; 97% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167075 bases at least Q40
                Consensus quality: 168567 bases at least Q30
                Consensus quality: 169221 bases at least Q20
                Insert size: 160000; agarose-fp
                Insert size: 169592; sum-of-contigs
                Quality coverage: 8.0 in Q20 bases; agarose-fp
```

```
            Quality coverage: 7.5 in Q20 bas.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     6481: contig of 6481 bp in length
          *     6482     6581: gap of 100 bp
          *     6582     6995: contig of 414 bp in length
          *     6996     7095: gap of 100 bp
          *     7096    47880: contig of 40785 bp in length
          *    47881    47980: gap of 100 bp
          *    47981    49647: contig of 1667 bp in length
          *    49648    49747: gap of 100 bp
          *    49748    54536: contig of 4789 bp in length
          *    54537    54636: gap of 100 bp
          *    54637    58724: contig of 4088 bp in length
          *    58725    58824: gap of 100 bp
          *    58825    63405: contig of 4581 bp in length
          *    63406    63505: gap of 100 bp
          *    63506    70593: contig of 7088 bp in length
          *    70594    70693: gap of 100 bp
          *    70694    83758: contig of 13065 bp in length
          *    83759    83858: gap of 100 bp
          *    83859    94549: contig of 10691 bp in length
          *    94550    94649: gap of 100 bp
          *    94650   109171: contig of 14522 bp in length
          *   109172   109271: gap of 100 bp
          *   109272   126036: contig of 16765 bp in length
          *   126037   126136: gap of 100 bp
          *   126137   170138: contig of 44002 bp in length
          *   170139   170238: gap of 100 bp
          *   170239   170892: contig of 654 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170892
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-110F1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .6481
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             6482. .6581
                     /estimated_length=100
     misc_feature    6582. .6995
                     /note="assembly_fragment"
     gap             6996. .7095
                     /estimated_length=100
     misc_feature    7096. .47880
                     /note="assembly_fragment"
     gap             47881. .47980
                     /estimated_length=100
     misc_feature    47981. .49647
                     /note="assembly_fragment"
     gap             49648. .49747
                     /estimated_length=100
     misc_feature    49748. .54536
                     /note="assembly_fragment"
     gap             54537. .54636
                     /estimated_length=100
     misc_feature    54637. .58724
                     /note="assembly_fragment"
     gap             58725. .58824
                     /estimated_length=100
     misc_feature    58825. .63405
                     /note="assembly_fragment"
     gap             63406. .63505
                     /estimated_length=100
     misc_feature    63506. .70593
                     /note="assembly_fragment"
     gap             70594. .70693
                     /estimated_length=100
     misc_feature    70694. .83758
                     /note="assembly_fragment"
     gap             83759. .83858
                     /estimated_length=100
     misc_feature    83859. .94549
                     /note="assembly_fragment"
     gap             94550. .94649
                     /estimated_length=100
     misc_feature    94650. .109171
                     /note="assembly_fragment"
     gap             109172. .109271
                     /estimated_length=100
     misc_feature    109272. .126036
                     /note="assembly_fragment"
     gap             126037. .126136
                     /estimated_length=100
     misc_feature    126137. .170138
                     /note="assembly_fragment"
     gap             170139. .170238
                     /estimated_length=100
     misc_feature    170239. .170892
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             47.3%;  Score 236.8;  DB 14;  Length 170892;
  Best Local Similarity   73.0%;  Pred. No. 4.2e-56;
  Matches  371;  Conservative    0;  Mismatches  107;  Indels   30;  Gaps    4;

Qy          5 GGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCC 64
              | ||||||||  ||  ||||||||||||||||  ||||| |||||  | |||||  ||| |||
Db     120205 GTCTCACTCTCTCACCCAGGCTGGAGTGCAATGGCGCGATCTTGGCTCACTGCAAACTCT 120146

Qy         65 ATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCT 124
                ||||||||||||||| ||||||  |||||||||||| ||||||||||||
Db     120145 GCCTCCTGGGTTCAAATGATTCTCGTGCCTCAGCCTCCCAAGTAGCTGGGA--------- 120095

Qy        125 CTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAG 184
                |  | || |||| |||||| ||||||||||||||||   ||| ||||||||||||
Db     120094 ------TTACAGGTACACGCCACCAAGCCCAGCTAATTTTTT---TGTATTTTTTGGTAG 120044

Qy        185 AGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCC----------TTTTAGAT 233
              | ||||  ||||||||||||||||||  |        | |            | |
Db     120043 AAACAGAATTTCACCATGTTGGCCAGGCCAGTCTCAAACTCCTGACCTCAAGTGATCTGC 119984

Qy        234 TCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC 293
              |||   ||| |  | |  |   || |  ||    | | || |||   ||   |
Db     119983 CTGCCTTGGCCTCCCACAGTGCTAGGATTACAGGTGTGAGCTACCGTGTCCAGCCTATGC 119924

Qy        294 ACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGA 353
              | || | || || || || || || ||||||| |||||| ||| |||| || ||||||||
Db     119923 ATATCTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCACCCAGGCTGGA 119864

Qy        354 GTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCC 413
              || |||||| | ||||||||||||||||||||||||  || | ||||||||||||||||||
Db     119863 GTGCAGTGGCGGGATCTCGGCTCACTGCAAGCTCCGTCTCCCGGGTTCACGCCATTCTCC 119804

Qy        414 CGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTT 473
              |||||||||||| |||||||||||||| ||||||| |||| |||||| ||||| |||
Db     119803 TGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCCGCCACTACGCCCGGCTAA-TTT 119745

Qy        474 TTTGTATTTTTAATAAAGACGGGGTTTC 501
              |||||||||||| || ||||||||||||
Db     119744 TTTGTATTTTTAGTAGAGACGGGGTTTC 119717
```

```
RESULT 13
AC099811/c
LOCUS       AC099811              174902 bp    DNA     linear   PRI 07-FEB-2003
DEFINITION  Homo sapiens chromosome 17, clone RP11-358B23, complete sequence.
ACCESSION   AC099811
VERSION     AC099811.7  GI:28269667
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 174902)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-358B23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174902)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 174902)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 174902)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 7, 2003 this sequence version replaced gi:27552629.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12016
                Center clone name: 358_B_23
            --------------.
FEATURES             Location/Qualifiers
     source          1. .174902
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-358B23"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          716. .726
                     /note="<30 qual SNGL region"
     unsure          1097. .1125
                     /note="<30 qual SNGL region"
     unsure          1132. .1136
                     /note="<30 qual SNGL region"
     unsure          1159. .1163
                     /note="<30 qual SNGL region"
     repeat_region   complement(1779. .1880)
                     /rpt_family="MIR"
     repeat_region   2702. .2725
                     /rpt_family="(CAAAA)n"
     repeat_region   3108. .3284
                     /rpt_family="(CA)n"
     repeat_region   complement(4282. .4533)
                     /rpt_family="AluJo"
     repeat_region   complement(4569. .4651)
                     /rpt_family="MIR"
     repeat_region   complement(5226. .5270)
                     /rpt_family="MER5B"
     repeat_region   5421. .5490
                     /rpt_family="MIR"
     repeat_region   6164. .6323
                     /rpt_family="AluSg/x"
     repeat_region   complement(6707. .6996)
                     /rpt_family="AluSg"
     repeat_region   complement(7002. .7122)
                     /rpt_family="AluJb"
     repeat_region   complement(7125. .7257)
                     /rpt_family="MER5A"
     repeat_region   7138. .7247
                     /rpt_family="MER5B"
     repeat_region   8182. .8372
                     /rpt_family="MIR"
     repeat_region   complement(8432. .8587)
                     /rpt_family="MIR3"
```

```
     repeat_region   8766. .9069
                     /rpt_family="AluSx"
     repeat_region   9081. .9200
                     /rpt_family="L1ME3"
     repeat_region   9348. .9652
                     /rpt_family="AluSc"
     repeat_region   complement(11210. .11443)
                     /rpt_family="AluY"
     repeat_region   12226. .12257
                     /rpt_family="(TA)n"
     repeat_region   complement(12258. .12566)
                     /rpt_family="AluSx"
     repeat_region   complement(12567. .12866)
                     /rpt_family="AluJb"
     repeat_region   13359. .13584
                     /rpt_family="MIR"
     repeat_region   complement(13951. .14075)
                     /rpt_family="AluJb"
     repeat_region   15039. .15095
                     /rpt_family="L1MC3"
     repeat_region   15096. .15381
                     /rpt_family="AluSx"
     repeat_region   15382. .15524
                     /rpt_family="L1MC3"
     repeat_region   complement(15525. .15826)
                     /rpt_family="AluSx"
     repeat_region   15827. .15915
                     /rpt_family="L1MC3"
     repeat_region   15916. .16226
                     /rpt_family="AluSq"
     repeat_region   16227. .16271
                     /rpt_family="L1MC3"
     repeat_region   16273. .16386
                     /rpt_family="AluY"
     repeat_region   16392. .16721
                     /rpt_family="AluY"
     repeat_region   16722. .16939
                     /rpt_family="AluY"
     repeat_region   16943. .17252
                     /rpt_family="L1MC3"
     repeat_region   17253. .17507
                     /rpt_family="AluSx"
     repeat_region   17508. .17535
                     /rpt_family="L1MC3"
     repeat_region   complement(17536. .17847)
                     /rpt_family="AluSx"
     repeat_region   17848. .18159
                     /rpt_family="L1MC3"
     repeat_region   18161. .18209
                     /rpt_family="AluJo"
     repeat_region   18212. .18506
                     /rpt_family="AluSc"
     repeat_region   18507. .18738
                     /rpt_family="AluY"
     repeat_region   18739. .18922
                     /rpt_family="(GAAA)n"

  Query Match             46.9%;  Score 235.2;  DB 8;  Length 174902;
  Best Local Similarity   73.0%;  Pred. No. 1.2e-55;
  Matches  365;  Conservative    0;  Mismatches  113;  Indels   22;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | |||||||||||| |||||||||||| ||| |||| |  ||||||| |||||||| ||
Db      30569 GAGTCTCACTCTGTCACCCAGGCTGGAATGCAGTGGTGCAATCTCAGCTCACTACAAGCT 30510

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||||||||||    |||||| |||||||| ||| ||||||||||||||||
Db      30509 CCGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGGCTCCCAAGTAGCTGGGACT----- 30455

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                    | ||| | | |||||||  ||| ||||||||||||     | ||||| ||
Db      30454 ----------ACAGGCGCCCGCCACCACACCCGGCTAATTTTTTG-----TATTTTTAGT 30410

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||||| |||||||||||||| |||||| | |||  |    |    | | | ||  ||
Db      30409 GGAGACAGGGTTTCACCATGTTAGCCAGGAT-GGTCTCAATCTCCTGACCTCCTGATCCG 30351

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              |||||    | || ||    |    |         ||   |   | | |      |||
Db      30350 CCTGCCTCGTCCTCCCAAAGTGCTGGGATTACAAGTGTCAGCCACCGCACCCAGCCTATA 30291

Qy        303 ATTAT-TCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTG 361
              |  || | || || || ||||||| ||||||  || ||||||| |||||||||| |||||
Db      30290 AAAATATTTTTTTTTTTTGAGACGGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTG 30231

Qy        362 GTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAG 421
              |||| |||||||||||||||| |||| ||| | ||||||||||||||||||| |||||||
Db      30230 GTGCAATCTCGGCTCACTGCAAACTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAG 30171

Qy        422 CCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATT 481
              |||||| |||||||||||||   |||||||||||||||||| ||||||||||||||||
Db      30170 CCTCCCGAGTAGCTGGGACTACCGGCGCCTGCCACCACGCCCAGCTAATTTTTTTGTATC 30111

Qy        482 TTTAATAAAGACGGGGTTTC 501
              |||| || ||||||||||||
Db      30110 TTTAGTAGAGACGGGGTTTC 30091


RESULT 14
AL357936
LOCUS       AL357936               107529 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-524G24 on chromosome 9 Contains
            the 5' end of two variants of the DAB2IP gene for DAB2 interacting
            protein (AF9Q34, DIP1/2, KIAA1743), a novel gene and two CpG
            islands, complete sequence.
ACCESSION   AL357936
VERSION     AL357936.25  GI:18252285
KEYWORDS    HTG; AF9Q34; CpG island; DAB2IP; DIP1/2; KIAA1743.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 107529)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 20, 2002 this sequence version replaced gi:18250750.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-524G24 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
```

```
FEATURES             Location/Qualifiers
     source          1. .107529
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-524G24"
                     /clone_lib="RPCI-11.2"
     misc_feature    2000
                     /note="Clone_right_end: RP11-162D16"
     misc_feature    8212. .8295
                     /note="Sequence from reads from a short insert library
                     derived from a single pUC clone. Restriction digest data
                     confirm the assembly. Sequence from overlapping clone
                     RP11-508N19 (AL353609). Assembly confirmed by restriction
                     digest."
     misc_feature    9570. .9633
                     /note="Sequence from reads from a short insert library
                     derived from a single pUC clone. Restriction digest data
                     confirm the assembly. Sequence from overlapping clone
                     RP11-508N19 (AL353609). Assembly confirmed by restriction
                     digest."
     misc_feature    9940. .9982
                     /note="Sequence from reads from a short insert library
                     derived from a single pUC clone. Restriction digest data
                     confirm the assembly. Sequence from overlapping clone
                     RP11-508N19 (AL353609). Assembly confirmed by restriction
                     digest."
     misc_feature    9983. .10001
                     /note="Sequence from overlapping clone RP11-508N19
                     (AL353609). Assembly confirmed by restriction digest."
     gene            join(44539. .44884,AL365274.17:50805. .50908,
                     AL365274.17:71452. .71585,AL365274.17:129140. .129293,
                     AL365274.17:131025. .131123,AL365274.17:132012. .132075,
                     AL365274.17:156611. .157655)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-002"
     mRNA            join(44539. .44884,AL365274.17:50805. .50908,
                     AL365274.17:71452. .71585,AL365274.17:129140. .129293,
                     AL365274.17:131025. .131123,AL365274.17:132012. .132075,
                     AL365274.17:156611. .157655)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-002"
                     /product="DAB2 interacting protein"
                     /note="match: ESTs: BF565256.1
                     match: cDNAs: AK096391.1"
     gene            join(44758. .44884,AL365274.17:50805. .50908,
                     AL365274.17:71452. .71585,AL365274.17:129140. .129293,
                     AL365274.17:131025. .131123,AL365274.17:132012. .>132455)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-005"
     mRNA            join(44758. .44884,AL365274.17:50805. .50908,
                     AL365274.17:71452. .71585,AL365274.17:129140. .129293,
                     AL365274.17:131025. .131123,AL365274.17:132012. .>132455)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-005"
                     /product="DAB2 interacting protein"
                     /note="match: ESTs: BM545676.1 BM546615.1"
     CDS             join(44845. .44884,AL365274.17:50805. .50908,
                     AL365274.17:71452. .71585,AL365274.17:129140. .129293,
                     AL365274.17:131025. .131123,AL365274.17:132012. .>132455)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-005"
                     /standard_name="OTTHUMP00000022029"
                     /codon_start=1
                     /product="DAB2 interacting protein"
                     /protein_id="CAI15112.1"
                     /db_xref="GI:55959205"
                     /db_xref="InterPro:IPR000008"
                     /db_xref="InterPro:IPR001849"
                     /db_xref="UniProt/TrEMBL:Q5T4Q1"
                     /translation="MEPDSLLDQDDSYESPQERPGSRRSLPGSLSEKSPSMEPSAATP
                     FRVTGFLSRRLKGSIKRTKSQPKLDRNHSFRHILPGFRSAAAAAADNERSHLMPRLKE
                     SRSHESLLSPSSAVEALDLSMEEEVVIKPVHSSILGQDYCFEVTTSSGSKCFSCRSAA
                     ERDKWMENLRRAVHPNKDNSRRVEHILKLWVIEAKDLPAKKKYLCELCLDDVLYARTT
                     GKLKTDNVFWGEHFEFHNLPPLRTVTVHLYRETDKKKKKERNSYLGLVSLPAASVAGR
                     QFVEKWYPVVTPNPKGGKGPGPMIRIKARYQTITILPMEMYKEFAEHIT"
     gene            complement(join(52547. .53068,53731. .53870))
                     /locus_tag="RP11-524G24.2-001"
     mRNA            complement(join(52547. .53068,53731. .53870))
                     /locus_tag="RP11-524G24.2-001"
                     /note="match: ESTs: BM715539.1"
     misc_feature    75919. .76068
                     /note="Sequence from overlapping clone RP11-508N19
                     (AL353609). Assembly confirmed by restriction digest."
     gene            join(75924. .76282,AL365274.17:50805. .50909)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-003"
     mRNA            join(75924. .76282,AL365274.17:50805. .50909)
                     /gene="DAB2IP"
                     /locus_tag="RP11-298A17.1-003"
                     /product="DAB2 interacting protein"
                     /note="match: ESTs: BI489865.1"
     misc_feature    77213. .77257
                     /note="Sequence from overlapping clone RP11-508N19
                     (AL353609). Assembly confirmed by restriction digest."
     misc_feature    105530
                     /note="Clone_left_end: RP11-298A17"
ORIGIN

  Query Match             46.9%;  Score 234.8;  DB 8;  Length 107529;
  Best Local Similarity   70.1%;  Pred. No. 1.5e-55;
  Matches  352;  Conservative    0;  Mismatches  132;  Indels   18;  Gaps    2;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| ||||||||||||||||||||||| |||||| |||||||| ||||| ||| ||
Db      10020 GAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCAGCTCACTGCAAGCT 10079

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              |   ||||||||||||    |||||| |||||||||||| ||||||||||||||||
Db      10080 CTGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACT----- 10134

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTGGT 182
                        | |||||   |||||| |||  |||||||||||||| | ||||||| ||
Db      10135 ----------ACAGGCACCTGCCACCAGGCCTGGCTAATTTTTTGTATTTTTTTTTTAGT 10184

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||||| |||||||||||| || ||| |||       || | |      |   | |
Db      10185 AGAGACAGGGTTTCACCATGTTAGCTAGGATGGTCTCGATCTCCTGACCTTGTGCTCCGC 10244

Qy        243 CCTGCTCTATTCTTCCCCTTTCT---AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||  |||       |   | ||   | | |||     ||   ||  ||    |  |||
Db      10245 CCATCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCACACCCGGCCTCTTT 10304

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
               || || | ||| || || ||||||  ||||||||||| |||| || ||||||||||| |||
Db      10305 TTTCTTTTACTTTTTTTTTTTGAGATGGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAG 10364

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              | |  | |||| |||||||||||| |||   || ||| |||||  | ||||||| |||||
Db      10365 TAGCACAATCTTGGCTCACTGCAATCTCTGTCTCCTGAGTTCAAACAATTCTCCTGCCTC 10424

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||  | |||||||||||| || | |    | |||| | ||||  | || |||||||||
Db      10425 AGCCTTTCGAGTAGCTGGGATTACAAGTATGTACCACAATGCCCGCCGAAGTTTTTTGTA 10484

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||| || ||||||||||||
Db      10485 TTTTTAGTAGAGACGGGGTTTC 10506


RESULT 15
AL356115
LOCUS       AL356115              150151 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-486O22 on chromosome 10 Contains
            the 3' end of a novel gene (KIAA1128, FLJ14262, FLJ25809) , the 5'
```

```
            end of a Siah-interacting protein (SIP) (calcyclin binding protein)
            pseudogene, ribosomal protein S3A pseudogene 5 (RPS3AP5) and two
            CpG islands, complete sequence.
ACCESSION   AL356115
VERSION     AL356115.9  GI:9795038
KEYWORDS    HTG; calcyclin; FLJ14262; FLJ25809; KIAA1128; RPS3A5;
            Siah-interacting; SIP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 150151)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 11, 2000 this sequence version replaced gi:9713038.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            RP11-486O22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .150151
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-486O22"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-486O22"
     misc_feature    53062
                     /note="Clone_left_end: RP11-470J18"
     gene            join(complement(AC074321.10:101526. .>102204),
                     complement(AC074321.10:100180. .100376),
                     complement(AC074321.10:56134. .56224),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:35288. .35483),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     76190. .76274,89764. .94832)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-002"
     mRNA            join(complement(AC074321.10:101526. .>102204),
                     complement(AC074321.10:100180. .100376),
                     complement(AC074321.10:56134. .56224),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:35288. .35483),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     76190. .76274,89764. .94832)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-002"
                     /product="KIAA1128"
                     /note="match: cDNAs: AB032954 AB032954.1 AK024324.1"
     gene            join(complement(AC074321.10:80053. .80142),
                     complement(AC074321.10:56134. .56224),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     89764. .90720)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-003"
     mRNA            join(complement(AC074321.10:80053. .80142),
                     complement(AC074321.10:56134. .56224),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     89764. .90720)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-003"
                     /product="KIAA1128"
                     /note="match: cDNAs: AK098675.1"
     gene            join(complement(AC074321.10:21690. .21778),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     89764. .90185)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-007"
     mRNA            join(complement(AC074321.10:21690. .21778),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     89764. .90185)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-007"
                     /product="KIAA1128"
                     /note="match: ESTs: BI461360.1 BI461611.1"
     gene            join(complement(AC074321.10:35288. .35452),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     76190. .76274,83609. .83648,89764. .90017)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-008"
     mRNA            join(complement(AC074321.10:35288. .35452),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     76190. .76274,83609. .83648,89764. .90017)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-008"
                     /product="KIAA1128"
                     /note="match: ESTs: BQ722236.1"
     gene            join(complement(AC074321.10:21690. .21778),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),
                     complement(AC074321.10:849. .963),53890. .53979,
                     89764. .90011)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-005"
     mRNA            join(complement(AC074321.10:21690. .21778),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),
                     complement(AC074321.10:849. .963),53890. .53979,
                     89764. .90011)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-005"
                     /product="KIAA1128"
                     /note="match: ESTs: BG719330.1 BQ429698.1"
     gene            join(complement(AC074321.10:48769. .48813),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:35288. .35483),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
```

```
                     89764. .89976)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-006"
     mRNA            join(complement(AC074321.10:48769. .48813),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:35288. .35483),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     89764. .89976)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-006"
                     /product="KIAA1128"
                     /note="match: ESTs: BQ440427.1"
     CDS             join(complement(AC074321.10:101526. .>102204),
                     complement(AC074321.10:100180. .100376),
                     complement(AC074321.10:56134. .56224),
                     complement(AC074321.10:48102. .48264),
                     complement(AC074321.10:35288. .35483),
                     complement(AC074321.10:9979. .10062),
                     complement(AC074321.10:3537. .3623),53890. .53979,
                     76190. .76274,89764. .89858)
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-006"
                     /standard_name="OTTHUMP00000020001"
                     /note="match: proteins: Q9D6G6 Q9H1N6 Q9H2S0 Q9H7U1
                     Q9ULU1"
                     /codon_start=1
                     /product="KIAA1128"
                     /protein_id="CAI40902.1"
                     /db_xref="GI:57209021"
                     /db_xref="UniProt/TrEMBL:Q9ULU1"
                     /translation="AVDLTKPYQNQQLSIRVPLRSSMLTRNSRQPEVLNGNEHLGYGF
                     NRPYAAGGKKLALPNGPGVTSTLGYRMVHPSLLKSSRSPFSGTMTVDGNKNSPADTCV
                     EEDATVLAKDRAANKDQELIENESYRTKNNQTMKHDAKMRYLSDDVDDISLSSLSSSD
                     KNDLSEDFSDDFIDIEDSNRTRITPEEMSLKEEKHENGPPQDMFDSPKENEKAFSKTD
                     EWIDISVSDRSECTKHTSGNNLVSPDTDYRAGSSFELSPSDSSDGTYMWDEEGLEPIG
                     NVHPVGSYESSEMNSIDILNNLESCDLEDDDLMLDVDLPEDAPLENVECDNMNRFDRP
                     DRNVRQPQEGFWKRPPQRWSGQEHYHLSHPDHYHHHGKSDLSRGSPYRESPLGHFESY
                     GGMPFFQAQKMFVDVPENTVILDEMTLRHMVQDCTAVKTQLLKLKRLLHQHDGSGSLH
                     DIQLSLPSSPEPEDGDKVYKNEDLLNEIKQLKDEIKKKDEKIQLLELQLATQHICHQK
                     CKEEKCTYADKYTQTPWRRIPGGYSAPSFSPWQGSFQGIPRTVPPHRRQTSSTTAFQQ
                     PSQTHRSHPGKTNKATTYRGPQ"
     polyA_signal    90701. .90706
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-003"
     polyA_site      90722
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-002"
     polyA_signal    91100. .91105
                     /gene="KIAA1128"
                     /locus_tag="RP11-486O22.1-002"

  Query Match             46.7%;  Score 234.2;  DB 8;  Length 150151;
  Best Local Similarity   72.9%;  Pred. No. 2.3e-55;
  Matches  364;  Conservative    0;  Mismatches  113;  Indels   22;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||| |||||||||||||||| ||||||  ||||| |  |||||||| ||
Db     137980 GAGTCTGGCTCTGTAGCCCAGGCTGGAGTGCAGTGGCGCAATCTCGGCGCACTACAAGCT 138039

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||| ||||||| |||| |  |||||| |||||||||||| | ||||||||||||
Db     138040 CCACCTCCTGGTTTCACACCATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGACT----- 138094

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | ||| | | |||||| |||| |||||||||||    ||| ||||||  ||
Db     138095 ----------ACAGGCGCCCGCCACCACGCCCGGCTAATTTTT----TGTATTTTTTAGT 138140

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||  || ||||||| |||| |||||| | |||  |   |   | |  |  |  ||
Db     138141 GGAGACGGGGTTTCACCGTGTTAGCCAGGAT-GGTCTCGATCTCCTGACCTCGTGATCTG 138199

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              |||||  |  ||  |||  |  ||  |  |  |  |  ||  |||  ||  |
Db     138200 CCTGCCTCAGCCT--CCCAAAGTGCTGGGATTACAGGCGTGACTAAGGTTCATTTTGTCT 138257

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              | || || || || ||||||||||||||| ||| ||||||| |||||||||| ||| ||
Db     138258 GATTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGCGG 138317

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              || |||||||||||||| ||||||| ||| | |||||||||||||||||| |||| |||
Db     138318 CGCCATCTCGGCTCACTGGAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTTAGC 138377

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              |||||  |||||||||| || ||||||| |||||||||||| ||||||||||||||||||
Db     138378 CTCCCGGGTAGCTGGGAGTACAGGCGCCCGCCACCACGCCCGGCTAATTTTTTTGTATTT 138437

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || |||||| |||||
Db     138438 TTAGTAGAGACGGTGTTTC 138456


Search completed: February  9, 2006, 18:20:26
Job time : 2701.09 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 344.959 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 501
Sequence: 1 acgggctcactctgtcgcc.....tttaataaagacggggtttc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	12174	12	ADI35082	Adi35082 Human PLA
	2	501	100.0	12174	12	ADJ09983	Adj09983 Human pho
	3	501	100.0	13612	6	ABK47376	Abk47376 Human Pho
	4	429.2	85.7	3375	3	AAA35101	Aaa35101 Human ade
	5	429.2	85.7	3375	3	AAF21223	Aaf21223 Human low
	6	429.2	85.7	3375	10	ABZ96917	Abz96917 Human nuc
	7	429.2	85.7	3375	11	ABD20766	Abd20766 Human pul
	8	429.2	85.7	6771	3	AAA35103	Aaa35103 Human ade
	9	429.2	85.7	6771	3	AAF21225	Aaf21225 Human low
	10	429.2	85.7	6771	10	ABZ96919	Abz96919 Human nuc
	11	429.2	85.7	6771	11	ABD20768	Abd20768 Human pul
c	12	231.6	46.2	35973	6	ABK13076	Abk13076 Human amy
c	13	231.4	46.2	602	4	AAL06681	Aal06681 Human rep
c	14	231.4	46.2	602	4	ABA08020	Aba08020 Human ova
	15	228.8	45.7	127508	13	ABD33171	Abd33171 Human can
	16	228.4	45.6	73100	14	ADZ42280	Adz42280 Human end
c	17	228.4	45.6	109906	6	ABK94411	Abk94411 DNA encod
c	18	228.4	45.6	109906	12	ADL08112	Adl08112 Human gen
c	19	227.6	45.4	10331	4	AAK79915	Aak79915 Human imm
c	20	227.6	45.4	10331	4	AAL03344	Aal03344 Human rep
	21	224.2	44.8	129042	11	ACN44674	Acn44674 Human gen
c	22	224	44.7	14175	4	AAS27814	Aas27814 DNA encod
	23	224	44.7	14175	4	AAK78859	Aak78859 Human imm
c	24	224	44.7	14175	10	ADB94617	Adb94617 Novel hum
	25	223	44.5	260209	6	ABS56564	Abs56564 Human SUL
	26	223	44.5	260209	12	ADN16204	Adn16204 Human sul
	27	222.8	44.5	73995	11	ACN43986	Acn43986 Human gen
	28	222.6	44.4	3067	10	ADE09851	Ade09851 Novel DNA
c	29	222.6	44.4	36534	11	ADL82795	Adl82795 Human sem
	30	222.2	44.4	24167	5	ABA16132	Aba16132 Human ner
	31	222	44.3	33769	5	ABA82622	Aba82622 Human HBM
	32	222	44.3	33769	6	ABK22781	Abk22781 Human hig
	33	222	44.3	33769	8	ACC45363	Acc45363 Human HBM
	34	222	44.3	33769	10	ADB98063	Adb98063 HBM-relat
	35	222	44.3	33769	10	ADE82432	Ade82432 Human DNA
	36	222	44.3	33769	13	ADR16926	Adr16926 BAC clone
	37	222	44.3	33769	13	ADR47577	Adr47577 BAC clone
	38	222	44.3	33769	14	AEB69306	Aeb69306 Human Hig
c	39	222	44.3	137046	13	ABD32936	Abd32936 Human can
c	40	222	44.3	156843	11	ACN44786	Acn44786 Human gen
	41	221.8	44.3	12595	4	AAS42100	Aas42100 Genomic s
c	42	221	44.1	611	4	AAK71483	Aak71483 Human imm
c	43	220.4	44.0	890	4	AAH99337	Aah99337 Human pro
	44	220.4	44.0	3884	4	AAH98627	Aah98627 Human EST
	45	220.2	44.0	32152	4	ABA08132	Aba08132 Human ova

ALIGNMENTS

RESULT 1
ADI35082
ID ADI35082 standard; DNA; 12174 BP.
XX
AC ADI35082;
XX
DT 22-APR-2004 (first entry)
XX
DE Human PLA2G1B nucleotide sequence.
XX
KW PLA2G1B ; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

```
CC   site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC   subject, where the presence of the polymorphic variation indicates a
CC   predisposition to fat deposition in the subject. The polymorphic
CC   variation is a guanine at position 7328 or thymine at position 9182 of
CC   the present sequence. The method is useful for diagnosing a
CC   predisposition to fat deposition or leanness in a subject, and
CC   consequently for diagnosing a predisposition to non-insulin dependent
CC   diabetes mellitus (NIDDM) in a subject and conditions such as
CC   hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC   hypercoagulability, or microalbuminuria, which can lead to early
CC   prescription of preventive measures. The present sequence represents a
CC   human PLA2G1B nucleotide sequence.
XX
SQ   Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

  Query Match             100.0%;  Score 501;  DB 12;  Length 12174;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-133;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5000 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 5059

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5060 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 5119

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5120 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 5179

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5180 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 5239

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5240 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 5299

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5300 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 5359

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5360 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 5419

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5420 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 5479

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       5480 TTTTAATAAAGACGGGGTTTC 5500


RESULT 2
ADJ09983
ID   ADJ09983 standard; DNA; 12174 BP.
XX
AC   ADJ09983;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW   human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW   PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW   appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW   non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW   hypertension; antidiabetic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(436,c)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(839,a)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4050,a)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(4689,t)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6282,a)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6358,c)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(6653,t)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7256,t)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7300,a)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7301,c)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(7328,g)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(8062,c)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(9182,t)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphisms"
FT   variation       replace(11649,c)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphisms"
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   P-PSDB; ADJ09984.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
```

```
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ   Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 501;  DB 12;  Length 12174;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-133;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5000 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 5059

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5060 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 5119

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5120 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 5179

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5180 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 5239

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5240 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 5299

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5300 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 5359

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5360 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 5419

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5420 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 5479

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       5480 TTTTAATAAAGACGGGGTTTC 5500


RESULT 3
ABK47376
ID   ABK47376 standard; DNA; 13612 BP.
XX
AC   ABK47376;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW   Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(3845,A)
FT                   /*tag=  a
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(3968,A)
FT                   /*tag=  b
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4053. .9613
FT                   /*tag=  c
FT                   /product= "Pla2G1B"
FT   exon            4053. .4086
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4087. .5785
FT                   /*tag=  e
FT                   /number= 1
FT   exon            5786. .5945
FT                   /*tag=  f
FT                   /number= 2
FT   intron          5946. .6744
FT                   /*tag=  g
FT                   /number= 2
FT   variation       replace(6060,A)
FT                   /*tag=  h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   exon            6745. .6873
FT                   /*tag=  i
FT                   /number= 3
FT   variation       replace(6844,A)
FT                   /*tag=  j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   intron          6874. .9489
FT                   /*tag=  k
FT                   /number= 3
FT   exon            9490. .9613
FT                   /*tag=  l
FT                   /number= 4
FT   variation       replace(9531,A)
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
```

```
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 501;  DB 6;  Length 13612;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-133;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4516 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 4575

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4576 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 4635

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4636 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 4695

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4696 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 4755

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4756 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 4815

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4816 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 4875

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4876 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 4935

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4936 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 4995

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       4996 TTTTAATAAAGACGGGGTTTC 5016


RESULT 4
AAA35101
ID   AAA35101 standard; DNA; 3375 BP.
XX
AC   AAA35101;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2790.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US017712.
XX
PR   03-AUG-1998;   98US-0095212P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstruction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers.
XX
PS   Disclosure; Page 1056-1057; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an antisense
CC   oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC   nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC   impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of the
CC   ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC   sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC   from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC   AAA33992) are specifically claimed ONs from the present invention. N.B.
CC   Sequences given in the disclosure of the present invention do not match
CC   up with their corresponding SEQ ID NO: sequences given in the sequence
CC   listing
XX
SQ   Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 429.2;  DB 3;  Length 3375;
  Best Local Similarity   97.4%;  Pred. No. 9.8e-113;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  ||||||||||||||||||||
Db        996 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1054

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db       1055 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112
```

```
Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTTG 1172

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1232

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||    |||
Db       1233 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1292

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1293 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1352

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1412

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1413 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 1471

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       1472 TTTTTAATAAAGACGGGGTTTC 1493
```

RESULT 5
AAF21223
ID AAF21223 standard; DNA; 3375 BP.
XX
AC AAF21223;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2790.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX
PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
XX
PS Disclosure; Page 1135-1136; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

  Query Match   85.7%; Score 429.2; DB 3; Length 3375;
  Best Local Similarity   97.4%; Pred. No. 9.8e-113;
  Matches 489; Conservative   0; Mismatches   8; Indels   5; Gaps   5;

```
Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  |||||||||||||||||||||
Db        996 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1054

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||
Db       1055 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1172

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1232

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||    |||
Db       1233 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1292

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1293 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1352

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1412

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1413 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 1471

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       1472 TTTTTAATAAAGACGGGGTTTC 1493
```

```
RESULT 6
ABZ96917
ID   ABZ96917 standard; DNA; 3375 BP.
XX
AC   ABZ96917;
XX
DT   17-OCT-2003  (first entry)
XX
DE   Human nucleic acid sequence.
XX
KW   Human; antisense; lung dysfunction; nasal airway dysfunction;
KW   antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW   antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW   antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW   adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW   lung inflammation; respiratory disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285308-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013135.
XX
PR   24-APR-2001; 2001US-0286137P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-229219/22.
XX
PT   Pharmaceutical composition for treating ailments associated with impaired
PT   respiration, has oligo(s) antisense to specific gene(s) or its
PT   corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT   ubiquinone.
XX
PS   Disclosure; SEQ ID NO 12159; 872pp; English.
XX
CC   The invention relates to a novel pharmaceutical composition, which has a
CC   first active agent comprising an oligonucleotide antisense to the
CC   initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC   5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC   junctions of genes encoding a polypeptide associated with lung and/or
CC   nasal airway dysfunction and a second active agent comprising an
CC   antiinflammatory steroid and ubiquinone. A composition of the invention
CC   has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC   immunosuppressive, and cytostatic activity. The composition may have a
CC   use in antisense gene therapy. The composition is useful for treating or
CC   preventing a respiratory, lung or malignant disease or condition, also
CC   for enhancing the prophylactic or therapeutic respiratory effect of an
CC   antiinflammatory steroid in a subject, for reducing or depleting levels
CC   of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC   receptor, producing bronchodilation, increasing levels of ubiquinone or
CC   lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC   lung inflammation, lung allergies, or a respiratory disease or condition.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 429.2;  DB 10;  Length 3375;
  Best Local Similarity   97.4%;  Pred. No. 9.8e-113;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||| |||||||||||||||  |||||||||||||||||||||
Db        996 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1054

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db       1055 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1172

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1232

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              |||||||||||||||||||||||| |||||||||||||||||||||||     |||
Db       1233 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1292

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1293 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1352

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1412

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1413 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 1471

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       1472 TTTTTAATAAAGACGGGGTTTC 1493


RESULT 7
ABD20766
ID   ABD20766 standard; DNA; 3375 BP.
XX
AC   ABD20766;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human pulmonary and inflammatory target DNA #377.
XX
KW   Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW   respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW   surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW   analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW   beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW   respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW   emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW   pulmonary transplantation rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285309-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013143.
XX
PR   24-APR-2001; 2001US-0286036P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-093058/08.
XX
PT   Pharmaceutical composition for treating asthma, has antisense
PT   oligonucleotide containing less percentage of adenosine, targeted to
PT   nucleic acids associated with lung airway or lung dysfunction, and
PT   bronchodilating agent.
XX
```

PS Claim 15; SEQ ID NO 12159; 763pp; English.
XX
CC This invention describes a novel composition (a) a first active agent,
CC comprising oligonucleotides, effective for alleviating
CC bronchoconstriction, respiratory tract inflammation, allergies and
CC reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC surfactant depletion or hyposecretion, when administered to a mammal. The
CC oligonucleotides are derived from a gene encoding or regulating
CC expression of a target polypeptide associated with lung airway or lung
CC dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC The invention also describes a kit, that comprises: (a) a delivery
CC device, in separate containers, (b) the oligonucleotides, (c)
CC instructions for adding a carrier and for use of the kit. The composition
CC of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC beta-adrenergic agonist. The composition is useful for preventing or
CC treating a respiratory, lung or malignant disease. The administered
CC composition comprises oligo and is administered to reduce the production
CC or availability, or to increase the degradation of the target mRNA or to
CC reduce the amount of target polypeptide present in the lungs. The
CC pulmonary obstruction, and/or bronchoconstriction and/or lung
CC inflammation, allergies and/or surfactant hypoproduction are associated
CC with a disease or condition such as pulmonary vasoconstriction,
CC inflammation, allergies, asthma, impeded respiration, respiratory
CC distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC transplantation rejection, pulmonary infections, bronchitis or cancer.
CC The reduced adenosine content of the anti-sense oligos corresponding to
CC thymidines present in the target RNA serves to prevent the breakdown of
CC the oligonucleotides into products that free adenosine into the system
CC e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC prevent any unwanted effects due to it
XX
SQ Sequence 3375 BP; 837 A; 866 C; 764 G; 908 T; 0 U; 0 Other;

Query Match 85.7%; Score 429.2; DB 11; Length 3375;
Best Local Similarity 97.4%; Pred. No. 9.8e-113;
Matches 489; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

```
Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||| |||||||||||||||||  |||||||||||||||||||||
Db        996 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1054

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||| ||||||||
Db       1055 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1112

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1172

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1232

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||||||||||||||||||||||||| |||||||||||||||||||||||||    |||
Db       1233 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1292

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       1293 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1352

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1412

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1413 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 1471

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
```

Db 1472 TTTTTAATAAAGACGGGGTTTC 1493

RESULT 8
AAA35103
ID AAA35103 standard; DNA; 6771 BP.
XX
AC AAA35103;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2792.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 1058-1059; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing
XX
SQ Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

```
  Query Match             85.7%;  Score 429.2;  DB 3;  Length 6771;
  Best Local Similarity   97.4%;  Pred. No. 1.3e-112;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  |||||||||||||||||||||
Db       1558 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1616

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db       1617 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1734

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1794

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||||||||||||||||||||||| |||||||||||||||||||||||     |||
Db       1795 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1854

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db       1855 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1914

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1974

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1975 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 2033

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       2034 TTTTTAATAAAGACGGGGTTTC 2055


RESULT 9
AAF21225
ID   AAF21225 standard; DNA; 6771 BP.
XX
AC   AAF21225;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2792.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US008020.
XX
PR   06-APR-1999;   99US-0127958P.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not trigger
PT   adenosine receptors during metabolism, useful e.g. for treating cancers
PT   and respiratory obstructions.
XX
PS   Disclosure; Page 1137-1139; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC   surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 429.2;  DB 3;  Length 6771;
  Best Local Similarity   97.4%;  Pred. No. 1.3e-112;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  |||||||||||||||||||||
Db       1558 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1616

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db       1617 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1734

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1794

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              ||||||||||||||||||||||| |||||||||||||||||||||||     |||
Db       1795 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1854

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db       1855 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1914

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
```

```
Db       1915 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1974

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1975 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 2033

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||||
Db       2034 TTTTTAATAAAGACGGGGTTTC 2055
```

RESULT 10
ABZ96919
ID ABZ96919 standard; DNA; 6771 BP.
XX
AC ABZ96919;
XX
DT 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.
XX
PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12161; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match   85.7%; Score 429.2; DB 10; Length 6771;
  Best Local Similarity 97.4%; Pred. No. 1.3e-112;
  Matches 489; Conservative   0; Mismatches   8; Indels   5; Gaps   5;

```
Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  ||||||||||||||||||||
Db       1558 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1616

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||
Db       1617 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1734

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1794

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              |||||||||||||||||||||||| |||||||||||||||||||||||||    |||
Db       1795 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1854

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1855 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1914

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1974

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 479
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1975 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 2033

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||||
Db       2034 TTTTTAATAAAGACGGGGTTTC 2055
```

RESULT 11
ABD20768
ID ABD20768 standard; DNA; 6771 BP.
XX
AC ABD20768;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human pulmonary and inflammatory target DNA #379.
XX
KW Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW pulmonary transplantation rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200285309-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013143.
XX
PR 24-APR-2001; 2001US-0286036P.
XX

```
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-093058/08.
XX
PT   Pharmaceutical composition for treating asthma, has antisense
PT   oligonucleotide containing less percentage of adenosine, targeted to
PT   nucleic acids associated with lung airway or lung dysfunction, and
PT   bronchodilating agent.
XX
PS   Claim 15; SEQ ID NO 12161; 763pp; English.
XX
CC   This invention describes a novel composition (a) a first active agent,
CC   comprising oligonucleotides, effective for alleviating
CC   bronchoconstriction, respiratory tract inflammation, allergies and
CC   reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC   surfactant depletion or hyposecretion, when administered to a mammal. The
CC   oligonucleotides are derived from a gene encoding or regulating
CC   expression of a target polypeptide associated with lung airway or lung
CC   dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC   The invention also describes a kit, that comprises: (a) a delivery
CC   device, in separate containers, (b) the oligonucleotides, (c)
CC   instructions for adding a carrier and for use of the kit. The composition
CC   of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC   analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC   beta-adrenergic agonist. The composition is useful for preventing or
CC   treating a respiratory, lung or malignant disease. The administered
CC   composition comprises oligo and is administered to reduce the production
CC   or availability, or to increase the degradation of the target mRNA or to
CC   reduce the amount of target polypeptide present in the lungs. The
CC   pulmonary obstruction, and/or bronchoconstriction and/or lung
CC   inflammation, allergies and/or surfactant hypoproduction are associated
CC   with a disease or condition such as pulmonary vasoconstriction,
CC   inflammation, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC   transplantation rejection, pulmonary infections, bronchitis or cancer.
CC   The reduced adenosine content of the anti-sense oligos corresponding to
CC   thymidines present in the target RNA serves to prevent the breakdown of
CC   the oligonucleotides into products that free adenosine into the system
CC   e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC   prevent any unwanted effects due to it
XX
SQ   Sequence 6771 BP; 1815 A; 1570 C; 1439 G; 1947 T; 0 U; 0 Other;

  Query Match             85.7%;  Score 429.2;  DB 11;  Length 6771;
  Best Local Similarity   97.4%;  Pred. No. 1.3e-112;
  Matches  489;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    5;

Qy          1 ACGGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAAC 60
              |||||||||||||||||| |||||||||||||||||  ||||||||||||||||||||
Db       1558 ACGGGGCTCACTCTGTCG-CCAGGCTGGAGTGCGGTGGGCTGATCTCAGATCACTACAAC 1616

Qy         61 CTCCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGT 120
              |||||||||| |||||||||||||||||||||||||||||||||||| ||||||||
Db       1617 CTCCATCTCCT-GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCT-GGACTTGT 1674

Qy        121 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 AGCTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTG 1734

Qy        181 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GTAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTC 1794

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCT-AATGCAGTATCCAGTTTCCTTACTTATCACATTT 299
              |||||||||||||||||||||||| |||||||||||||||||||||    |||
Db       1795 ATCCTGCTCTATTCTTCCCCTTTCTAAATGCAGTATCCAGTTTCCTTACTTATACACTTT 1854

Qy        300 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAG 359
              |||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||

Db       1855 ATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTGCCCAAGGCTGGAGTACAG 1914

Qy        360 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 TGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTC 1974

Qy        420 AGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTTGTA 479
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1975 AGCCTCCCCAGTAGCTGGGACTAAA-GCGCCTGCCACCACGCCCCGCTAATTTTTTTGTA 2033

Qy        480 TTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||||||
Db       2034 TTTTTAATAAAGACGGGGTTTC 2055


RESULT 12
ABK13076/c
ID   ABK13076 standard; DNA; 35973 BP.
XX
AC   ABK13076;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human amyloid beta precursor protein binding protein 1 gene.
XX
KW   Human; amyloid beta precursor protein binding protein 1; APPBP1; gene;
KW   Alzheimer's disease; transgenic animal; platelet aggregation;
KW   single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(3436,G)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3474,C)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             4085. .31975
FT                   /*tag=  c
FT                   /product= "APPBP1"
FT                   /note= "Amyloid beta precursor protein binding protein 1"
FT   variation       replace(11314,G)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18123,G)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18571,A)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21229,A)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21312,A)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24300,A)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26093,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29334,C)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31499,T)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31582,G)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
XX
```

```
PN   WO200202820-A1.
XX
PD   10-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-US020951.
XX
PR   30-JUN-2000; 2000US-0215511P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Chew A,  Choi JY,  Kazemi A,  Koshy B,  Sausker EA;
PI   Stephens CJ;
XX
DR   WPI; 2002-164539/21.
DR   P-PSDB; AAU11030.
XX
PT   Amyloid beta precursor protein binding protein 159 kD (APPBP1) gene
PT   polymorphic variants, useful e.g. in studying the expression and function
PT   of APPBP1 and screening candidate drugs for treating Alzheimer's disease.
XX
PS   Claim 1; Fig 1; 104pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a sequence
CC   which is a polymorphic variant of a reference sequence for the amyloid
CC   beta precursor protein binding protein 1, 59kD (APPBP1) protein or its
CC   fragment. The polymorphic variants are useful in studying the expression
CC   and function of APPBP1, in expressing APPBP1 protein for use in screening
CC   for candidate drugs to treat diseases related to APPBP1 activity, in
CC   studying the effect of the variation on the biological activity of
CC   APPBP1, and the binding affinity of candidate drugs targeting APPBP1 for
CC   the treatment of disorders such as Alzheimer's disease. The haplotyping
CC   methods are useful in validating APPBP1 as a candidate target for
CC   treating a specific condition or disease predicted to be associated with
CC   APPBP1 activity, or in the design of clinical trials of candidate drugs
CC   for treating a specific condition or disease associated with APPBP1
CC   activity. The transgenic animals are useful for studying expression of
CC   the APPBP1 isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against APPBP1 protein, and for testing the efficacy of
CC   therapeutic agents and compounds for disorders related to platelet
CC   aggregation in a biological system. The present sequence represents the
CC   human APPBP1 gene as described in the method of the invention
XX
SQ   Sequence 35973 BP; 9909 A; 7327 C; 7959 G; 10766 T; 0 U; 12 Other;

  Query Match             46.2%;  Score 231.6;  DB 6;  Length 35973;
  Best Local Similarity   72.4%;  Pred. No. 1.6e-55;
  Matches  349;  Conservative    0;  Mismatches  114;  Indels   19;  Gaps    3;

Qy         20 CCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTGGGTTCAA 79
              ||||||||||||||| |||||  |||||  | |||||  |||||||| |||||||||||||
Db      22424 CCAGGCTGGAGTGCAGTGGCACGATCTTGGCTCACTGCAACCTCCGTCTCCTGGGTTCAA 22365

Qy         80 ATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGCTGGCA 139
                 |||| | |||||||||||||  | | ||||||||||||                 | ||
Db      22364 GCGATTCCCCTGCCTCAGCCTCCCGAATAGCTGGGACT--------------ACAGGTG 22320

Qy        140 CACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTTTCACC 199
              | ||||||||  ||  ||||||||||||    ||||||| |||||||||| |||||||
Db      22319 CGCACCACCACACCTGGCTAATTTTTTGT---ATTTTTTTAGTAGAGACAGGGTTTCACC 22263

Qy        200 ATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCC 259
              |||||||||||| | |||       | | | |  || |  ||  || |        || ||
Db      22262 ATGTTGGCCAGGAT-GGTCTTGATCTTCTGACCTTGTGATCCACCCGTCTCGGCCTCCCG 22204

Qy        260 CTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTAT 319
                |    |         ||   |   |    |     ||   |  || || || || |
Db      22203 AAGTGGTGGGATTACAGGCGTGAGCCACCACGCCCGGCCTAAGTTCTTTTTTTTTTTTT 22144

Qy        320 TGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACT 379
              |||||||||||||  || ||||||| |||||||||| |||||||||||||||| |||||||
Db      22143 TGAGACAGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGTGCGATCTCCGCTCACT 22084

Qy        380 GCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGA 439
              |||||||||| ||| | |||||||||| ||||||| |||||||||||||  ||||||||||
Db      22083 GCAAGCTCCGCCTCCCGGGTTCACGCTATTCTCCTGCCTCAGCCTCCCGGGTAGCTGGGA 22024

Qy        440 CTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
              ||| |||||||||||||||||||| |||||||||||||||||||| || |||| |||||
Db      22023 CTACAGGCGCCTGCCACCACGCCCGGCTAATTTTTTTGTATTTTTAGTAGAGACAGGGTT 21964

Qy        500 TC 501
              ||
Db      21963 TC 21962


RESULT 13
AAL06681/c
ID   AAL06681 standard; DNA; 602 BP.
XX
AC   AAL06681;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9369.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
```

PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 9369; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 602 BP; 183 A; 152 C; 175 G; 92 T; 0 U; 0 Other;

Query Match 46.2%; Score 231.4; DB 4; Length 602;
Best Local Similarity 72.3%; Pred. No. 3.7e-56;
Matches 362; Conservative 0; Mismatches 121; Indels 18; Gaps 4;

```
Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  || ||||||| |||||||||||| |||||| |||||| | ||||| ||||||
Db        571 GAGTCTTGCTTTGTCGCCTAGGCTGGAGTGCAGTGGCGCGATCTCGGGTCACTGCAACCT 512

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||| |||||||    ||||||| |||||||||||| | ||||||||||||||
Db        511 CCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 457

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGT-GTGTTTTTTTTGG 181
                    | ||| | |||||||| |||| ||||||||||||  | |||||||| |
Db        456 ----------ACAGGCGCCCACCACCACGCCCGGCTAATTTTTTGTATTTTTTTTTTTAG 407

Qy        182 TAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCA 241
              |||||||||| ||||||| |||| |||||| |||       || |             |
Db        406 TAGAGACAGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCGGACCTCGTGGTCCG 347

Qy        242 TCCTGCTC-TATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              ||  |||    ||       |          |    | |   |  | ||       | |||
Db        346 CCCGCCTCGGCCTCCGAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCTTTT 287

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              || || || || || || |||||||||||||||| | |||| || |||||||||| ||||
Db        286 TTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGTTCTGTCCCCCAGGCTGGAGTGCAGT 227

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              || |||||||||||||||||||||||||||| |||||||||| ||||||||| ||||||
```

```
Db        226 GGCGCGATCTCGGCTCACTGCAAGCTCCACCTCCTGGGTTCACACCATTCTCCTGCCTCA 167

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||| ||||||||| |||| |||||||||||||| | |  ||||| ||||||||||
Db        166 GCCTCCCGAGTAGCTGGCACTACAGGCGCCTGCCACCATGACTGGCTAA-TTTTTTGTAT 108

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              ||||| || ||||||||||||
Db        107 TTTTAGTAGAGACGGGGTTTC 87


RESULT 14
ABA08020/c
ID   ABA08020 standard; DNA; 602 BP.
XX
AC   ABA08020;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human ovarian and breast cancer associated polynucleotide SEQ ID NO 815.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155325-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001345.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
```

```
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488786/53.
XX
PT   New isolated ovarian and/or breast cancer related nucleic acids and
PT   polypeptides, useful for diagnosing, treating and/or preventing human
PT   diseases and disorders, particularly ovarian and/or breast cancer.
XX
PS   Disclosure; SEQ ID NO 815; 577pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA07454-ABA08224) and proteins
CC   (ABB10743-ABB10980) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 602 BP; 183 A; 152 C; 175 G; 92 T; 0 U; 0 Other;

  Query Match            46.2%;  Score 231.4;  DB 4;  Length 602;
  Best Local Similarity   72.3%;  Pred. No. 3.7e-56;
  Matches  362;  Conservative    0;  Mismatches  121;  Indels   18;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  || ||||||| |||||||||||| |||||| |||||| | ||||| ||||||
Db        571 GAGTCTTGCTTTGTCGCCTAGGCTGGAGTGCAGTGGCGCGATCTCGGGTCACTGCAACCT 512

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||| |||||||    |||||| |||||||||||| | ||||||||||||
Db        511 CCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 457

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGT-GTGTTTTTTTTGG 181
                  | ||| | |||||||| |||| |||||||||||||  | ||||||| |
Db        456 ----------ACAGGCGCCCACCACCACGCCCGGCTAATTTTTTGTATTTTTTTTTTAG 407

Qy        182 TAGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCA 241
              |||||||||| ||||||| |||| |||||| |||       || |               |
Db        406 TAGAGACAGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCGGACCTCGTGGTCCG 347

Qy        242 TCCTGCTC-TATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
               ||  |||    ||      |       |   |   |   | | |     | |||
Db        346 CCCGCCTCGGCCTCCGAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCTTTT 287

Qy        301 TTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGT 360
              || || || || || || |||||||||||||||| | |||| || |||||||||| ||||
Db        286 TTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGTTCTGTCCCCAGGCTGGAGTGCAGT 227

Qy        361 GGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCA 420
              || |||||||||||||||||||||||||||||| |||||||||| ||||||||| ||||||
Db        226 GGCGCGATCTCGGCTCACTGCAAGCTCCACCTCCTGGGTTCACACCATTCTCCTGCCTCA 167

Qy        421 GCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
              ||||||| |||||||| |||| |||||||||||||| | |  ||||| ||||||||||
Db        166 GCCTCCCGAGTAGCTGGCACTACAGGCGCCTGCCACCATGACTGGCTAA-TTTTTTGTAT 108

Qy        481 TTTTAATAAAGACGGGGTTTC 501
              ||||| || ||||||||||||
Db        107 TTTTAGTAGAGACGGGGTTTC 87


RESULT 15
ABD33171
ID   ABD33171 standard; DNA; 127508 BP.
XX
AC   ABD33171;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-024.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 158; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
```

```
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 127508 BP; 33168 A; 27130 C; 29586 G; 37624 T; 0 U; 0 Other;

  Query Match             45.7%;  Score 228.8;  DB 13;  Length 127508;
  Best Local Similarity   70.8%;  Pred. No. 1.7e-54;
  Matches  358;  Conservative    0;  Mismatches  122;  Indels   26;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| ||||||| |||||||||||||||| |||| | |||||| | ||||| ||| ||
Db     106050 GAGTCTCGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGCGATCTCGGCTCACTGCAAGCT 106109

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||| |||||||    |||||| |||||||||||| | ||||||||||||
Db     106110 CCGCCTCCCGGGTTCACCCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT----- 106164

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | ||| | |||||||||||||||||||||||||    |  | ||||| ||
Db     106165 ----------ACAGGCGCCCACCACCATGCCCAGCTAATTTTT----TTGTATTTTTAGT 106210

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTT-------TTAGATTC 235
              |||||||||| ||||||| |||| |||||| ||        || | |         |    |
Db     106211 AGAGACAGGGTTTCACCCTGTTAGCCAGGATGATCTCGATCTCCTGACCTCGGGATCCGC 106270

Qy        236 TCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCAC 295
              | || |       | |   | |      |   | ||   | |||              |
Db     106271 CCGTCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGTCCGGCCTTTTT 106330

Qy        296 ATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGT 355
              |   || || || || ||  | ||||||| |||||||||| ||||||| ||||||||||
Db     106331 TTGCCTTTTTTTTTTTTTTTGTTTTGAGACGGAGTCTTGCTCTGTCGCCCAGGCTGGAGT 106390

Qy        356 ACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCG 415
              |||||| || |||| |||||||||||||||||| ||| | |||||| ||||||||| |
Db     106391 GCAGTGGCGCCATCTTGGCTCACTGCAAGCTCCGCCTTCCGGGTTCGGGCCATTCTCCTG 106450

Qy        416 CCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTT 475
              |||||||||||| |||||||||||||| ||||||| |||||| ||||| ||||||||||
Db     106451 CCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCGCCACCGCGCCCGGCTAATTTTTT 106510

Qy        476 TGTATTTTTAATAAAGACGGGGTTTC 501
              |||||||||| || ||||||||||||
Db     106511 TGTATTTTTAGTAGAGACGGGGTTTC 106536


Search completed: February  9, 2006, 13:50:53
Job time : 346.959 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 2363.65 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 501
Sequence: 1 acgggctcactctgtcgcc.....tttaataaagacggggtttc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	213.2	42.6	810	9	BZ609826	BZ609826 WHABQ60TR
c 2	213.2	42.6	877	6	CD515864	CD515864 AGENCOURT
c 3	211.6	42.2	627	6	CD702127	CD702127 EST18651
4	211.4	42.2	759	1	AW976010	AW976010 EST388119
c 5	211.4	42.2	820	5	BU568843	BU568843 AGENCOURT
6	210.6	42.0	3775	4	CR627381	CR627381 Homo sapi
7	209.8	41.9	689	9	BZ605978	BZ605978 WHAAT77TF
c 8	209.6	41.8	642	5	BX496786	BX496786 DKFZp779L
9	209.4	41.8	631	7	CK002773	CK002773 AGENCOURT
c 10	208.4	41.6	623	6	CD723982	CD723982 oj28g12.y
11	208.2	41.6	891	5	BU174155	BU174155 AGENCOURT
12	208.2	41.6	1654	4	CR591254	CR591254 full-leng
13	207.4	41.4	777	8	DR761502	DR761502 HESC4_129
14	207.4	41.4	832	5	BU180590	BU180590 AGENCOURT
15	207.4	41.4	935	5	BU838401	BU838401 AGENCOURT
16	206.6	41.2	620	3	BP382151	BP382151 BP382151
17	206.6	41.2	771	5	BQ776208	BQ776208 UI-H-FH0-
18	206	41.1	671	3	BM722145	BM722145 UI-E-EO0-
c 19	206	41.1	782	5	BX644719	BX644719 DKFZp781N
20	205.6	41.0	520	5	BQ807308	BQ807308 NISC_kk01
21	204.4	40.8	618	3	BP382202	BP382202 BP382202
c 22	204.2	40.8	613	9	AQ419352	AQ419352 RPCI-11-2
23	204.2	40.8	630	1	AI457389	AI457389 tl73c12.x
c 24	203.6	40.6	787	8	CX786428	CX786428 HESC3_60_
c 25	203	40.5	2128	4	CR925964	CR925964 Pongo pyg
c 26	201.6	40.2	908	5	BU171410	BU171410 AGENCOURT
27	201.4	40.2	685	2	BI193100	BI193100 602947363
28	201.4	40.2	1014	5	BQ709843	BQ709843 AGENCOURT
29	200.6	40.0	743	8	DR759591	DR759591 HESC4_105
c 30	200.2	40.0	714	8	CX872940	CX872940 HESC4_74_
c 31	199.2	39.8	990	1	AW600804	AW600804 N26-20 Hu
c 32	198.6	39.6	718	7	CN359768	CN359768 170004180
c 33	198.4	39.6	732	2	BG542372	BG542372 602571737
34	198	39.5	512	1	AL135377	AL135377 DKFZp762G
c 35	197.6	39.4	538	9	AQ585309	AQ585309 RPCI-11-4
c 36	197.6	39.4	877	9	AQ739838	AQ739838 HS_5505_A
c 37	197.4	39.4	942	9	AQ743787	AQ743787 HS_5501_A
c 38	197	39.3	621	9	AQ636660	AQ636660 RPCI-11-4
39	196.2	39.2	687	10	AG165009	AG165009 Pan trogl
c 40	195.8	39.1	910	5	BQ722917	BQ722917 AGENCOURT
c 41	195.6	39.0	586	1	AL701696	AL701696 DKFZp686D
42	195.4	39.0	576	5	BX508289	BX508289 DKFZp686L
43	195.4	39.0	666	9	AQ342859	AQ342859 RPCI11-12
c 44	195.4	39.0	1921	4	AF370377	AF370377 Homo sapi
c 45	195.2	39.0	493	2	BF826830	BF826830 MR2-HN003

ALIGNMENTS

RESULT 1
BZ609826
LOCUS BZ609826 810 bp DNA linear GSS 08-JUN-2003
DEFINITION WHABQ60TR Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-11I23, genomic survey sequence.
ACCESSION BZ609826
VERSION BZ609826.1 GI:31518387
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 810)
AUTHORS Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.
TITLE End-sequence profiling: Sequence-based analysis of aberrant genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
PUBMED 12788976
COMMENT Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-11I23"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library (MCF7_1)"
/note="Vector: pECBAC1; Site_1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."
ORIGIN

```
  Query Match             42.6%;  Score 213.2;  DB 9;  Length 810;
  Best Local Similarity   70.1%;  Pred. No. 6.4e-31;
  Matches  338;  Conservative    0;  Mismatches  123;  Indels   21;  Gaps    3;

Qy         20 CCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTGGGTTCAA 79
              |||||||||||||| |||| |||||||| | ||||| || |||||| ||||||||||||
Db        239 CCAGGCTGGAGTGCAGTGGTGTGATCTCGGCTCACTGCAGCCTCCACCTCCTGGGTTCAA 298

Qy         80 ATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGCTGGCA 139
              | |||||| |||||||||||| | ||||||||||||                | |||
Db        299 GTGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACT--------------ACAGGCC 343

Qy        140 CACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTTTCACC 199
              |    | |||| |||||||||||||||||     | |||||  ||||||  || |||||||
Db        344 CGTGCTACCAAGCCCAGCTAATTTTTTG-----TATTTTTAATAGAGATGGGGTTTCACC 398

Qy        200 ATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCC 259
              |||||||||||| |||  |    || | || |  |  |  ||  |||||  |  ||| ||
Db        399 ATGTTGGCCAGGATGGTCGCAATCTC-TTGACCTCGTGATCTACCTGCCTTGGTCTCCCA 457

Qy        260 CTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTAT 319
                  |    |                |  |  |          | || || || |
Db        458 AAGTGCTGGGATTACAGGCCCGAGGCACTGTGCCTGGCGTGCCTGGCCTTTTTTTTTTT 517

Qy        320 TGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACT 379
              |||||| |||||| ||| ||||||| |||||||||| |||||| ||||||||||||||||
Db        518 TGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGATCTCGGCTCACT 577

Qy        380 GCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGA 439
              |||||||||||| ||| |||||| ||||| ||| | |||  ||  || |||||||| |
Db        578 GCAAGCTCCACCTCCTGAGTTCACACCATTTTCCTGGCTCCACCCTCCGAGTAGCTGGAA 637

Qy        440 CTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
              ||| | |||||  |||||| |||  |||||||||||||||||||| || |||||||||
Db        638 CTACAAGCGCCCACCACCATGCCTGGCTAATTTTTTTGTATTTTTAGTAGAGACGGGGTT 697

Qy        500 TC 501
              ||
Db        698 TC 699
```

```
RESULT 2
CD515864/c
LOCUS       CD515864                877 bp    mRNA    linear   EST 06-JUN-2003
DEFINITION  AGENCOURT_14367770 NIH_MGC_181 Homo sapiens cDNA clone
            IMAGE:30397388 5', mRNA sequence.
ACCESSION   CD515864
VERSION     CD515864.1  GI:31447582
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM480  row: j  column: 21
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .877
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30397388"
                     /tissue_type="White Matter"
                     /dev_stage="Unknown"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_181"
                     /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV
                     (destroyed); Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.42 kb. Library was constructed by
                     (Invitrogen). Note: this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             42.6%;  Score 213.2;  DB 6;  Length 877;
  Best Local Similarity   70.2%;  Pred. No. 6.3e-31;
  Matches  355;  Conservative    0;  Mismatches  123;  Indels   28;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||||| ||||| ||||||||||||||||  ||| | |||||||| ||||| ||||||
Db        548 GAGTCTCACCCTGTCACCCAGGCTGGAGTGCAATGGGGCGATCTCAGCTCACTGCAACCT 489

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||| |||||||| |  |||||||||||||||  |||  |||||||| ||
Db        488 CCGCCTCCCGGGTTCACACGATTCTCTTGCCTCAGGTTCTGGAGTAGCTGAGA------- 436

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  |  | ||| |   ||||||||||||||||||||||||     | |||||  ||
Db        435 --------TTACAGGCTCCTGCCACCATGCCCAGCTAATTTTT------GTATTTTTAGT 390

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTG-----GGTGACCTCCCTTTTAGATTCTC 237
              |||||| || |||||||||||||||||||||     | |  |||  | ||  ||| |||
Db        389 AGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGATCTCCTGACCTTGTGATCCTC 330

Qy        238 CTCATCCTGC--TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCAC 295
              |      | ||   | |  | | |    |  | ||   | |||   | | |
Db        329 CCACCTCAGCCTCCCAAAGTGCTCGGATTACAGGCGTGAGCCACCGTGCCCAGCCTATTT 270

Qy        296 ATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGT 355
              ||||||  |  | |  ||||| |||||||||||||||| | |||| || ||||||||||
Db        269 GTTTATTTATTATTTATTTATTTTTGAGACAGAGTCTTGTTCTGTCACCCAGGCTGGAGT 210

Qy        356 ACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCG 415
              ||||||   |||||| |||||||||||| |||| ||| | ||||||| || ||| ||| |
Db        209 GCAGTGGCATGATCTCAGCTCACTGCAACCTCCTCCTTCCGGGTTCAAGCAATTATCCTG 150

Qy        416 CCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTT 475
              |||||||||||| ||||  ||||| || |||||    |||| || ||| |||||||||| |
Db        149 CCTCAGCCTCCCTAGTAAGTGGGATTACAGGCGTGCACCACTACACCCGGCTAATTTTCT 90

Qy        476 TGTATTTTTAATAAAGACGGGGTTTC 501
              |  ||||||| || ||||||||||||
Db         89 TTCATTTTTAGTAGAGACGGGGTTTC 64
```

```
RESULT 3
CD702127/c
LOCUS       CD702127                627 bp    mRNA    linear   EST 25-JUN-2003
DEFINITION  EST18651 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION   CD702127
VERSION     CD702127.1  GI:32232757
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
            Zeng,Y.-X.
  TITLE     Transcriptional Gene Expression Profile of Human Nasopharynx
```

```
  JOURNAL   Unpublished (2003)
COMMENT     Contact: YiXin Zeng
            Cancer Center
            Sun Yat-sen University
            651 DongFeng Road East, GuangZhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@gzsums.edu.cn.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="normal nasopharynx"
                     /clone_lib="human nasopharynx"
                     /note="ESTs generated from a normal nasopharynx cDNA
                     library from southern Chinese"
ORIGIN

  Query Match             42.2%;  Score 211.6;  DB 6;  Length 627;
  Best Local Similarity   69.8%;  Pred. No. 1.4e-30;
  Matches  359;  Conservative    0;  Mismatches  119;  Indels   36;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| |||||||  ||||||||||||||  ||||| ||||||| ||||| ||||||
Db        623 GAGTCTCGCTCTGTCAGCCAGGCTGGAGTGCCATGGCGCGATCTCAGCTCACTGCAACCT 564

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              |   ||||  |||||||   |||||| |||||||||| |   ||||||||||
Db        563 CTGCCTCCCAGGTTCAAGCCATTCTCCTGCCTCAGCCACCTGAGTAGCTGGGA------- 511

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | | ||  | | | |  |||||||||||||||       | |||||  ||
Db        510 --------TTACAGGTGCGCTCAAAGACACCCAGCTAATTTTT------GTATTTTTAGT 465

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG----------GTGACCTCCCTTTTAGA 232
              ||||||||| ||||||  ||||||||||||||          | |||||   ||
Db        464 AGAGACAGGGTTTCACTGTGTTGGCCAGGCTGGTCTCAAACCTCTAACCTCAGGTTATCC 405

Qy        233 TTCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTA- 291
                || ||   |  | | |              || |     |   |   | ||| || ||
Db        404 ACCTGCTTCGGCCTCACAGAGTGCTGGGATTACAGGAGTGAGCCACCGTGCCTGACCTAT 345

Qy        292 ----TCACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAG 347
                  |  || || || || ||  | || || ||||||| |||||| ||| ||||||| ||
Db        344 ACGTTTTCAGTTTTTTTTTTTTGTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAG 285

Qy        348 GCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCA 407
              |||||||| |||||| | ||||| |||||||||||||||||||| ||||||||||||||
Db        284 GCTGGAGTGCAGTGGCGTGATCTTGGCTCACTGCAAGCTCCACCTCCTGGGTTCACGCCA 225

Qy        408 TTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCT 467
              |||||| |||||||||||| |||||||||||||||| ||||||| |||||||| ||| |||
Db        224 TTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCCGCCACCACACCCGGCT 165

Qy        468 AATTTTTTTGTATTTTTAATAAAGACGGGGTTTC 501
              |||||||||||||||||| || | || |||||||
Db        164 AATTTTTTTGTATTTTTAGTAGACACAGGGTTTC 131


RESULT 4
AW976010
LOCUS       AW976010                759 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  EST388119 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW976010
VERSION     AW976010.1  GI:8167232
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 356
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .759
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGN"
                     /note="Vector: pBluescriptSKm"
ORIGIN

  Query Match             42.2%;  Score 211.4;  DB 1;  Length 759;
  Best Local Similarity   70.0%;  Pred. No. 1.4e-30;
  Matches  355;  Conservative    0;  Mismatches  121;  Indels   31;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | |||||||||| ||||||||||||||||| ||| || |||||| | ||||| ||| ||
Db        131 GAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGACGCGATCTCGGCTCACTGCAAGCT 190

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||| |||| |||||||     |||||| |||||||||||| | ||||||||||
Db        191 CCACCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGA------- 243

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | | ||||||  |||||| |||| |||| |||||       | ||||| |
Db        244 --------TTACAGGCACATGCCACCACGCCCGGCTACTTTTT------GTATTTTTAGC 289

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG---------GTGACCTCCCTTTTAGAT 233
              |||||| || |||||||| ||| |||||||||                |  |  |  |
Db        290 AGAGACGGGGTTTCACCATCTTGCCCAGGCTGGTCTTGAGCTCCTGACCTCGTGATCCAC 349

Qy        234 TCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC 293
              | ||||  ||| |   | | |    || |   | |     | || |     | |
Db        350 CCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTAAGCCACTGTGCCTGGCTTGAAT 409

Qy        294 ACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGA 353
                ||| || ||  | || || || ||| ||| |||||| ||| || |||| ||||||||
Db        410 TTTTTTTTTTTTTCTTTTTTTTTTTTTGGGACGGAGTCTCGCTCTGCCGCCCAGGCTGGA 469

Qy        354 GTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCC 413
              || || ||| || |||||||||||||||||||||||||||| | |||| ||  ||||||||
Db        470 GTGCAATGGCGCCATCTCGGCTCACTGCAAGCTCCACCTCCCGGGTCCAGACCATTCTCT 529

Qy        414 CGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTT 473
              |||||||||||| |||||||||||||| ||| ||| |||||||||||| ||||| |||
Db        530 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGAGCCCGCCACCACGCCCGGCTAA-TTT 588

Qy        474 TTTGTATTTTTAATAAAGACGGGGTTT 500
              |||||||||||  | ||||||||||||
Db        589 TTTGTATTTTTAGCAGAGACGGGGTTT 615


RESULT 5
BU568843/c
LOCUS       BU568843                820 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10400218 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6616116
            5', mRNA sequence.
ACCESSION   BU568843
VERSION     BU568843.1  GI:22919143
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2859  row: e  column: 12
            High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6616116"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_82"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
ORIGIN
```

Query Match 42.2%; Score 211.4; DB 5; Length 820;
Best Local Similarity 70.3%; Pred. No. 1.4e-30;
Matches 351; Conservative 0; Mismatches 121; Indels 27; Gaps 4;

```
Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||||||||||||||||||||||||||| |||| | |||||  | ||||| ||| ||
Db        568 GAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGGGCGATCTTGGCTCACTGCAAGCT 509

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||||||||||    |||||| ||||||||||||   ||||||||||||
Db        508 CCGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACT----- 454

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                      | ||| |   ||||   |||| ||||||||||||     | ||||| ||
Db        453 ----------ACAGGCGCCTGCCACTGCGCCCGGCTAATTTTTTG-----TATTTTTAGT 409

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||  || |||||||||||| |||||| |||       || | |  |  |       ||
Db        408 AGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTGAACTTGTGATCCAC 349

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              |   |||         ||      | ||| |     |      | |   | | |
Db        348 CTGCCTC------GGCCTCCCAAAGTGCTGGGATTACAGATGTGAGCCACCGCGCCCGGC 295

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              || || || || || ||||||  ||||||  || |||||||  |||||||||| ||||||
Db        294 CTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGG 235

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              |||||||  |||||||||||||||||| ||| |||||||| |||||||| ||||||||
Db        234 GGCGATCTTGGCTCACTGCAAGCTCCGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGC 175

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||||||||  |||||  ||||| ||||||||||||
```

```
Db        174 CTCCTGAGTAGCTGGGACTACAGGCGCCTGCCACTGCGCCCGGCTAA-TTTTTTGTATTT 116

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||| |||||||||
Db        115 TTAGTAGAGATGGGGTTTC 97


RESULT 6
CR627381
LOCUS       CR627381               3775 bp    mRNA    linear   HTC 22-SEP-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp762F237 (from clone DKFZp762F237).
ACCESSION   CR627381
VERSION     CR627381.1  GI:50949857
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3775)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Medigenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project.
            This clone (DKFZp762F237) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp762F237
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3775
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762F237"
                     /tissue_type="melanoma (MeWo cell line)"
                     /clone_lib="762 (synonym: hmel2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="hypothetical protein, N-terminus truncated"
     gene            1. .3775
                     /gene="DKFZp762F237"
     CDS             <13. .744
                     /gene="DKFZp762F237"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH10478.1"
                     /db_xref="GI:50949858"
                     /translation="YGRVNVVEALQEFWQMKQSRGADLKNGALVVYEMVPSNSPPYVC
                     YVTLPGGSCFGSFQFCPTKAEARRSAAKIALMNSVFNEHPSRRITDEFIEKSVSEALA
                     SFNGNREEADNPNTGIGAFRFMLESNKGKSMLEFQELMTVFQLLHWNGSLKAMRERQC
                     SRQEVLAHYSHRALDDDIRHQMALDWVSREQSVPGALSRELASTERELDEARLAGKEL
                     RFHKEKKDILVLAAGQLGNMHSSNC"
ORIGIN
```

Query Match 42.0%; Score 210.6; DB 4; Length 3775;
Best Local Similarity 71.5%; Pred. No. 1.5e-30;
Matches 357; Conservative 0; Mismatches 89; Indels 53; Gaps 4;

```
Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||| |||||| |||||| | ||||| ||||||
Db       2219 GAGTCTTGCTCTGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCT 2278

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  | |||| |  |||||| |||||||||||| | ||||||||||
```

```
Db       2279 CCGCCTCCCAGATTCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA------- 2331

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                |  | ||  | ||||||| ||||||||||||||||||     | ||||| ||
Db       2332 --------TTACAGGTGCCCACCACCACGCCCAGCTAATTTTTTG-----TATTTTTAGT 2378

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||| || |||||||||||| ||||||  ||                |||||  |
Db       2379 AGAGACGGGGTTTCACCATGTTAGCCAGGACGG-----------------TCTCCATCT 2420

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||||  ||  |  ||| | | |    ||                |  |    |||| |
Db       2421 CCTGACCTCGTGATCCACCTGCCTCGGC---------------CTCCCAAACTATTTGTG 2465

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                | || || || || ||||||| |||||| ||| ||||||| |||||||||| ||||||
Db       2466 TGTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 2525

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
               || |||||||||||||||||||||| ||| | |||||||||||| ||||| ||||||||
Db       2526 CGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCTCAGC 2585

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||| ||||||| |||  ||||||||||||||||||
Db       2586 CTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTT 2645

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db       2646 TTAGTAGAGACGGGGTTTC 2664
```

RESULT 7
BZ605978

```
LOCUS       BZ605978                689 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHAAT77TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-5N10, genomic survey sequence.
ACCESSION   BZ605978
VERSION     BZ605978.1  GI:31514539
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-5N10"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN
```

```
  Query Match             41.9%;  Score 209.8;  DB 9;  Length 689;
  Best Local Similarity   67.1%;  Pred. No. 3e-30;
  Matches  335;  Conservative    0;  Mismatches  147;  Indels   17;  Gaps    2;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||| | |||||||||||||| |||| |||||||| | ||||| ||||||
Db         89 GAGTCTTGCTCTGTTGTCCAGGCTGGAGTGCAGTGGTGTGATCTCGGCTCACTGCAACCT 148

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              |  ||||||||||||||| |||||||||||||||||||   ||||||||||||    |
Db        149 CTGCCTCCTGGGTTCAAATGATTCTCTTGCCTCAGCCTCCTGAGTAGCTGGGACTATAA- 207

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                                  |  ||  ||| ||  ||| |||||||||| | |  | ||||| ||
Db        208 --------------GAGCATGCCATCACACCCCGCTAATTTTTTTTTTTGTATTTTTAGT 253

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||||| ||||||  |||  |||||| |||      || | |              ||
Db        254 AGAGACAGGGTTTCACTGTGTAAGCCAGGATGGTCTCAATCTCCTGACCTCATGATCCAC 313

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||  |||      |  | ||      |  |     ||     |      || ||
Db        314 CCGCCTCGGCCTCCCAAAATGCTGGGATTACAGGCGTGAGCCACCACTCCTGGCCTTGTT 373

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                | || || || || |||||||| ||||||  |||| |||||  |||||||||| | ||||
Db        374 TCTTTTTTTTTTTTTTTTGAGACCGAGTCTCACTTTATCGCCCAGGCTGGAGTGCTGTGG 433

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
                 || | || ||||||||| || ||||| |||||||| ||  |||||| |||||| |
Db        434 CATAATTTTGGTTCACTGCAACCTTCACCTCCTGGGTTCAAGCAGTTCTCCTGCCTCAAC 493

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||| |||||| || | | ||  ||||||||||  | |||  ||||||||||
Db        494 CTCCTGAGTAACTGGGATTACAAGTGCGCGCCACCACGCCTGGTTAA--TTTTTGTATTT 551

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        552 TTAGTAGAGACGGGGTTTC 570
```

RESULT 8
BX496786/c

```
LOCUS       BX496786                642 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp779L1329_r1 779 (synonym: hncc1) Homo sapiens cDNA clone
            DKFZp779L1329 5', mRNA sequence.
ACCESSION   BX496786
VERSION     BX496786.1  GI:32012740
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     EST (Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
```

```
          This clone (DKFZp779L1329) is available at the RZPD in Berlin.
          Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
          Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp779L1329"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="DH10B"
                     /clone_lib="779 (synonym: hncc1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB"
ORIGIN

  Query Match             41.8%;  Score 209.6;  DB 5;  Length 642;
  Best Local Similarity   70.8%;  Pred. No. 3.3e-30;
  Matches  358;  Conservative    0;  Mismatches  121;  Indels   27;  Gaps    5;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||||||||||||||||||||||||||||| |||||  |||||||| ||||| ||| ||
Db        537 GAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGATCTCAGCTCACTGCAATCT 478

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  ||| ||    |||||| ||||||||||||   ||||||||||||
Db        477 CCGCCTCCCAGGTCCACGCCATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACT----- 423

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | |||||| ||||||||||||  ||||||||||     |  | ||||| ||
Db        422 ----------ACAGGCACCCACCACCATGCCTGGCTAATTTTT----TTGTATTTTTAGT 377

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGG-----CTGGGTGACCTCCCTTTTAGATTCTC 237
              |||  | |  |||||||||||| ||||||     || |    |||  | ||  ||  | |
Db        376 AGAAGCGGAGTTTCACCATGTTAGCCAGGATGGTCTCGANCTCCTGACCTTGTGANCCGC 317

Qy        238 CTCATCCTGC--TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCAC 295
              ||  | | ||   | |     |    |  | ||   | |||  | |  |  |
Db        316 CTGCTTCGGCCTCCCAAAGGGCTGGGATTACAGGCGTGAGCCACCGTGCCCAGCCTTTCT 257

Qy        296 ATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGT 355
              |||| || |  || ||  | ||   ||||| |||||||||| |||||||  ||||||||||
Db        256 GTTTTTTTTGTTTGTTTGTTTTTAAGAGACGGAGTCTTGCTCTGTCGCCCAGGCTGGAGT 197

Qy        356 ACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCG 415
              ||||||   |||||||||||||||| ||||||   || |||||||| ||||||||| |
Db        196 GCAGTGGCATGATCTCGGCTCACTGAAAGCTCTGTCTCCTGGGTTCATGCCATTCTCCTG 137

Qy        416 CCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTT 475
              |||||||||||| |||||||||||||| ||| |||  |||||||||| ||||| |||||
Db        136 CCTCAGCCTCCCAAGTAGCTGGGACTACAGGTGCCCACCACCACGCCCGGCTAA-TTTTT 78

Qy        476 TGTATTTTTAATAAAGACGGGGTTTC 501
              |||||||||| || || |||||||||
Db         77 TGTATTTTTAGTAGAGGCGGGGTTTC 52


RESULT 9
CK002773
LOCUS       CK002773                631 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_16398704 NIH_MGC_227 Homo sapiens cDNA clone
            IMAGE:30716650 5', mRNA sequence.
ACCESSION   CK002773
VERSION     CK002773.1  GI:38528807
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Professor Miklss Palkovits
             cDNA Library Preparation: Michael Brownstein / Ted Usdin
            Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM257  row: a  column: 11
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30716650"
                     /tissue_type="Bulk tissue from Human Spinal cord"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_227"
                     /note="Organ: brain/CNS; Vector: pDNR-LIB; Site_1: SfiI
                     (directional); Site_2: SfiI (directional); Library is
                     oligo-dT primed and directionally cloned.5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3'
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T)30N-1N-3. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected for >0.5kb with an
                     average insert size of 1.3kb Library created in the
                     laboratory of Jonathan Kuo and Ted Usdin."
ORIGIN

  Query Match             41.8%;  Score 209.4;  DB 7;  Length 631;
  Best Local Similarity   71.3%;  Pred. No. 3.6e-30;
  Matches  356;  Conservative    0;  Mismatches   91;  Indels   52;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||| |||||| |||||| |||||| | ||||| ||||||
Db        109 GAGTCTTGCTCTGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCT 168

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  | |||| |  |||||| |||||||||||| | ||||||||||
Db        169 CCGCCTCCCAGATTCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA------- 221

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                    | | ||  | |||||||| ||||||||||||||    |  | ||||| ||
Db        222 --------TTACAGGTGCCCACCACCACGCCCAGCTAATTTTT----TTGTATTTTTAGT 269

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||  || ||||||||||||| ||||||  ||                    |||||   |
Db        270 AGAGACGGGGTTTCACCATGTTAGCCAGGACGG------------------TCTCCATCT 311

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||||  ||  |  ||| | | |    ||                    |  |    |||| |
Db        312 CCTGACCTCGTGATCCACCTGCCTCGGC--------------CTCCCAAACTATTTGTG 356

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              | || || || || ||||||| |||||| ||| ||||||| |||||||||| ||||||
Db        357 TGTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 416

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              || |||||||||||||||||||||| ||| | |||||||||||| ||||| ||||||||
Db        417 CGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCTCAGC 476

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||| ||||||| |||  |||||||||||||||||
Db        477 CTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTT 536
```

```
Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        537 TTAGTAGAGACGGGGTTTC 555


RESULT 10
CD723982/c
LOCUS       CD723982                623 bp    mRNA    linear   EST 26-JUN-2003
DEFINITION  oj28g12.y1 Human lacrimal gland, unamplified: oj Homo sapiens cDNA
            clone oj28g12 5', mRNA sequence.
ACCESSION   CD723982
VERSION     CD723982.1  GI:32274836
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Ozyildirim,A.M., Wistow,G.J., Gao,J., Wang,J., Dickinson,D.P.,
            Frierson,H.F. Jr and Laurie,G.W.
  TITLE     The lacrimal gland transcriptome is an unusually rich source of
            rare and poorly characterized gene transcripts
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 46 (5), 1572-1580 (2005)
   PUBMED   15851553
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 28  row: g  column: 12
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="oj28g12"
                     /tissue_type="lacrimal gland"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human lacrimal gland, unamplified: oj"
                     /note="Organ: Eye; Vector: pCMVSport6; RNA was extracted
                     from 2 human lacrimal glands. A directionally cloned cDNA
                     library in the pCMVSPORT6 vector(Life Technologies) was
                     constructed at Bioserve Biotechnology (Laurel MD)
                     essentially following the protocols of the SuperScript
                     Plasmid System full details of which are contained in the
                     manufacturer's Instruction manual
                     (http://www.lifetech.com/). First strand synthesis was
                     carried out using a Not I primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. EST analysis
                     was performed on the unamplified library at the NIH
                     Intramural Sequencing Center (NISC)."
ORIGIN

  Query Match             41.6%;  Score 208.4;  DB 6;  Length 623;
  Best Local Similarity   68.8%;  Pred. No. 5.6e-30;
  Matches  347;  Conservative    0;  Mismatches  146;  Indels   11;  Gaps    4;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||||||||||| |||||   ||||| | ||||| ||||||
Db        498 GAGTCTTGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAACCT 439

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  |||||||  || |||| ||||||||||||    |||||||||| |     |
Db        438 CCGCCTCCCAGGTTCAAGCAAATCTCCTGCCTCAGCCTCCTGTGTAGCTGGGATTACAGG 379

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
              | | |     ||   ||| | | | |         | |||||| | |  | ||||| ||
Db        378 CACGC-----GCCACCACGCCCGGCTAATTTTTTTTTTTTTTTTTTTTTGTATTTTTAGT 324

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGG--GTGACCTCCCTTTTAGATTCTCCTC 240
              ||||||||| ||||||||||||||||||||| |||      | || |        ||  ||| |
Db        323 AGAGACAGGATTTCACCATGTTGGCCAGGGTGGTCTCAATCTTCTGACCTCATGATCCAC 264

Qy        241 ATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTA 300
              |         |       |   | | |||   |  |   || | |  | |  ||||
Db        263 CCACCTTGGCTCCCAAAGTGCTGGGATTACAGGCACGCGCCACCATGCCCAGCTAATTTT 204

Qy        301 TTATTATTCTT---ATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTAC 357
              |   || || ||    |  || ||||||| |||| | ||| ||| | | ||||||||| | |
Db        203 TGTTTTTTGTTTGGTTGGTTTTTGAGACGGAGTTTCGCTCTGTTGTCCAGGCTGGAATGC 144

Qy        358 AGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCC 417
              |||||||||||||| |||||||||||||||| ||| | ||||||| |||| ||||| |||
Db        143 AGTGGTGCGATCTCAGCTCACTGCAAGCTCCGCCTCCCGGGTTCATGCCACTCTCCTGCC 84

Qy        418 TCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTG 477
              |||||||||| |||||||||||||| |||||||  ||||  || || | ||| |||||||
Db         83 TCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCACCACTGCGTCCGGTTAA-TTTTTTG 25

Qy        478 TATTTTTAATAAAGACGGGGTTTC 501
              |||||||| || ||||||||||||
Db         24 TATTTTTAGTATAGACGGGGTTTC 1


RESULT 11
BU174155
LOCUS       BU174155                891 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_8106214 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6265610
            5', mRNA sequence.
ACCESSION   BU174155
VERSION     BU174155.1  GI:22688139
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2433  row: a  column: 03
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6265610"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

```
  Query Match             41.6%;  Score 208.2;  DB 5;  Length 891;
  Best Local Similarity   68.3%;  Pred. No. 5.7e-30;
  Matches  341;  Conservative    0;  Mismatches  138;  Indels   20;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| |||||||||||||||||||||||| |||||| |||||| | ||||| ||| ||
Db         31 GAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCT 90

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||  |||||||    |||||| |||||||||||| ||||||||| |||||
Db         91 CCGCCTCTCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTCGGACT----- 145

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | |||    | |||||||||| |||||| |||    ||| |||||| ||
Db        146 ----------ACAGGCGTCCTCCACCATGCCTGGCTAATATTT----TGTATTTTTTAGT 191

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||| || ||||||| |||| |||||| |||      || | |                |
Db        192 AGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTAACCTGGTGATCCGC 251

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||  |||        |   | |        |      ||     |   | |
Db        252 CCGCCTCGGCCTCCCAAAGTATTGGGATTACCCGCGTGAGCCACCGCGCCCGGCTGCAAA 311

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                | |  || ||  |||||||||||||| ||| ||| ||| ||||||||||| ||||||
Db        312 TAATCTTTCTTTTTTTCTGAGACAGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 371

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
               |||||||||||||| ||| ||||| ||| | |||||||||||||||||| ||||||||
Db        372 CACGATCTCGGCTCACGGCACGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGC 431

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
               |||| |||||||||||||| ||| ||| |||||||||||| |||||  ||||||| |||
Db        432 TTCCCGAGTAGCTGGGACTACAGGGGCCCGCCACCACGCCCGGCTAA-CTTTTTGTGTTT 490

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        491 TTAGTAGAGACGGGGTTTC 509
```

RESULT 12
CR591254

LOCUS	CR591254 1654 bp mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CS0DF001YF08 of Fetal brain of Homo sapiens (human).
ACCESSION	CR591254
VERSION	CR591254.1 GI:50472061
KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1654)
AUTHORS	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue
REFERENCE	2 (bases 1 to 1654)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES	Location/Qualifiers
source	1. .1654 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DF001YF08" /tissue_type="Fetal brain" /plasmid="pCMVSPORT_6"

ORIGIN

```
  Query Match             41.6%;  Score 208.2;  DB 4;  Length 1654;
  Best Local Similarity   68.3%;  Pred. No. 5.1e-30;
  Matches  341;  Conservative    0;  Mismatches  138;  Indels   20;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||| |||||||||||||||||||||||| |||||| |||||| | ||||| ||| ||
Db        420 GAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCT 479

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  |||  |||||||    |||||| |||||||||||| ||||||||| |||||
Db        480 CCGCCTCTCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTCGGACT----- 534

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | |||    | |||||||||| |||||| |||    ||| |||||| ||
Db        535 ----------ACAGGCGTCCTCCACCATGCCTGGCTAATATTT----TGTATTTTTTAGT 580

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||| || ||||||| |||| |||||| |||      || | |                |
Db        581 AGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTAACCTGGTGATCCGC 640

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||  |||        |   | |        |      ||     |   | |
Db        641 CCGCCTCGGCCTCCCAAAGTATTGGGATTACCCGCGTGAGCCACCGCGCCCGGCTGCAAA 700

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                | |  || ||  |||||||||||||| ||| ||| ||| ||||||||||| ||||||
Db        701 TAATCTTTCTTTTTTTCTGAGACAGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 760

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
               |||||||||||||| ||| ||||| ||| | |||||||||||||||||| ||||||||
Db        761 CACGATCTCGGCTCACGGCACGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGC 820

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
               |||| |||||||||||||| ||| ||| |||||||||||| |||||  ||||||| |||
Db        821 TTCCCGAGTAGCTGGGACTACAGGGGCCCGCCACCACGCCCGGCTAA-CTTTTTGTGTTT 879

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        880 TTAGTAGAGACGGGGTTTC 898
```

RESULT 13
DR761502

LOCUS	DR761502 777 bp mRNA linear EST 21-JUL-2005
DEFINITION	HESC4_129_D09.g1_A037 NIH_MGC_262 Homo sapiens cDNA clone IMAGE:7971211 5', mRNA sequence.
ACCESSION	DR761502
VERSION	DR761502.1 GI:71054202
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 777)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs-r@mail.nih.gov Tissue Procurement: BresaGen, Inc.

```
            cDNA Library Preparation: Express Genomics, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
           University of Georgia
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM 17087  row: g  column: 17
           Seq primer: JENREV (CAGGAAACAGCTATGACC)
           High quality sequence stop: 777.
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:7971211"
                     /sex="male"
                     /tissue_type="embryonic stem"
                     /cell_type="human embryonic stem cells"
                     /cell_line="BG01"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="NIH_MGC_262"
                     /note="Vector: pExpress-1; Site_1: NotI; Site_2: EcoRV;
                     RNA obtained from human embryonic stem cells isolated from
                     the inner cell mass of blastocyst stage embryos and
                     differentiated to an early neural progenitor cell type.
                     Cell line id and NIH Registry designation is BG01.
                     Positive for Nestin and Musashi expression. Passage number
                     18. cDNA primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. This primary library
                     is non-normalized (normalized primary library is
                     NIH_MGC_259). It was constructed by Express Genomics
                     (Frederick, MD). Sequence ends have been trimmed to
                     exclude vector and regions below Phred quality 16. Note:
                     this is a Mammalian Gene Collection library."
ORIGIN

  Query Match            41.4%;  Score 207.4;  DB 8;  Length 777;
  Best Local Similarity  69.9%;  Pred. No. 8.4e-30;
  Matches  349;  Conservative    0;  Mismatches   96;  Indels   54;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||| |||||| |||||| | ||||| ||||||
Db        193 GAGTCTTGCTCTGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCT 252

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  | |||| |  |||||| |||||||||||| | ||||||||||
Db        253 CCGCCTCCCAGATTCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA------- 305

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | | ||  | |||||||| ||||||||||||||||     | ||||| ||
Db        306 --------TTACAGGTGCCCACCACCACGCCCAGCTAATTTTTTG-----TATTTTTAGT 352

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||  || ||||||||||||| ||||||  ||
Db        353 AGAGACGGGGTTTCACCATGTTAGCCAGGACGG-------------------------- 385

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
                  | | | || |||  |       |    || |  ||          ||  ||  |
Db        386 -------TCTCCATCTCCTGACCTCGTGATCCACCTGCCTCGGCCTCCCAAACTATTTGT 438

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
               |  || || || || ||||||| |||||| ||| ||||||| |||||||||| ||||||
Db        439 GTGTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 498

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
               || |||||||||||||||||||| ||| | |||||||||||| |||||| ||||||||
Db        499 CGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCTCAGC 558

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||| ||||||| |||  ||||||||||||||||||
Db        559 CTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTT 618
```

```
Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        619 TTAGTAGAGACGGGGTTTC 637


RESULT 14
BU180590
LOCUS       BU180590                 832 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7852598 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6140102
            5', mRNA sequence.
ACCESSION   BU180590
VERSION     BU180590.1  GI:22694574
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM13458  row: c  column: 15
           High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1. .832
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6140102"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
ORIGIN

  Query Match            41.4%;  Score 207.4;  DB 5;  Length 832;
  Best Local Similarity  69.9%;  Pred. No. 8.3e-30;
  Matches  349;  Conservative    0;  Mismatches   96;  Indels   54;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||| |||||| |||||| |||||| | ||||| ||||||
Db         60 GAGTCTTGCTCTGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCT 119

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  | |||| |  |||||| |||||||||||| | ||||||||||
Db        120 CCGCCTCCCAGATTCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA------- 172

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | | ||  | |||||||| ||||||||||||||||     | ||||| ||
Db        173 --------TTACAGGTGCCCACCACCACGCCCAGCTAATTTTTTG-----TATTTTTAGT 219

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              ||||||  || ||||||||||||| ||||||  ||
Db        220 AGAGACGGGGTTTCACCATGTTAGCCAGGACGG-------------------------- 252

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
                  | | | || |||  |       |    || |  ||          ||  ||  |
Db        253 -------TCTCCATCTCCTGACCTCGTGATCCACCTGCCTCGGCCTCCCAAACTATTTGT 305
```

```
Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              |  || || || || ||||||| |||||| ||| ||||||| |||||||||| ||||||
Db        306 GTGTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 365

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              || |||||||||||||||||||||||| ||| | |||||||||||| ||||| ||||||||
Db        366 CGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCTCAGC 425

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||| ||||||| |||  ||||||||||||||||||
Db        426 CTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTT 485

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        486 TTAGTAGAGACGGGGTTTC 504
```

```
RESULT 15
BU838401
LOCUS       BU838401                935 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_8122967 Lupski_dorsal_root_ganglion Homo sapiens cDNA
            clone IMAGE:6177002 5', mRNA sequence.
ACCESSION   BU838401
VERSION     BU838401.1  GI:24022796
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13554  row: e  column: 03
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6177002"
                     /sex="male"
                     /tissue_type="dorsal root ganglia"
                     /dev_stage="adult, 36 yr"
                     /lab_host="DH10B"
                     /clone_lib="Lupski_dorsal_root_ganglion"
                     /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                     NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                     Directionally cloned using the following adaptors:
                     5'-TCGACCCACGCGTCCG-3' and
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
                     1 kb for average insert length 1.7 kb. This is a primary
                     library, non-amplified. Library constructed by Life
                     Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
                     College of Medicine) and is available through Life
                     Technologies."
ORIGIN
```

```
  Query Match             41.4%;  Score 207.4;  DB 5;  Length 935;
  Best Local Similarity   69.9%;  Pred. No. 8.1e-30;
  Matches  349;  Conservative    0;  Mismatches   96;  Indels   54;  Gaps    3;

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
              | | ||  |||||||||||||||| |||||| |||||| |||||| | ||||| ||||||
Db        247 GAGTCTTGCTCTGTCGCCCAGGCTAGAGTGCAGTGGCGCGATCTCGGCTCACTGCAACCT 306

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
              ||  ||||  | |||| |  |||||| |||||||||||| | ||||||||||
Db        307 CCGCCTCCCAGATTCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA------- 359

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                  | | ||  | |||||||| ||||||||||||||||||     | ||||| ||
Db        360 --------TTACAGGTGCCCACCACCACGCCCAGCTAATTTTTTG-----TATTTTTAGT 406

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
              |||||| || |||||||||||| ||||||  ||
Db        407 AGAGACGGGGTTTCACCATGTTAGCCAGGACGG-------------------------- 439

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
                  | | | || |||  |       |    || |  ||          ||  || |
Db        440 -------TCTCCATCTCCTGACCTCGTGATCCACCTGCCTCGGCCTCCCAAACTATTTGT 492

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
              |  || || || || ||||||| |||||| ||| ||||||| |||||||||| ||||||
Db        493 GTGTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 552

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              || |||||||||||||||||||||||| ||| | |||||||||||| ||||| ||||||||
Db        553 CGCTATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCACTCTCCTGCCTCAGC 612

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||  |||||||||||||| ||||||| ||||||| |||  ||||||||||||||||||
Db        613 CTCCTTAGTAGCTGGGACTACAGGCGCCCGCCACCATGCCTGGCTAATTTTTTTGTATTT 672

Qy        483 TTAATAAAGACGGGGTTTC 501
              ||| || ||||||||||||
Db        673 TTAGTAGAGACGGGGTTTC 691
```

```
Search completed: February  9, 2006, 18:10:14
Job time : 2367.65 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:28:05 ; Search time 10772.2 Seconds
(without alignments)
10559.003 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 1998.6
Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1998.6	100.0	8368	8	AY438977	AY438977 Homo sapi
	2	1998.6	100.0	13612	6	AX377239	AX377239 Sequence
c	3	1998.6	100.0	122302	8	AC003982	AC003982 Homo sapi
c	4	1998.6	100.0	220384	14	AC078926	AC078926 Homo sapi
	5	1811.2	90.6	189729	14	AC073930	AC073930 Homo sapi
c	6	462.6	23.1	178887	14	AC068981	AC068981 Homo sapi
	7	462.6	23.1	190162	8	AC009244	AC009244 Homo sapi
	8	460.6	23.0	123192	14	AP000589	AP000589 Homo sapi
	9	460.4	23.0	150266	8	AC022408	AC022408 Homo sapi
	10	459.4	23.0	167996	8	AC021753	AC021753 Homo sapi
	11	457.4	22.9	137693	8	AC003689	AC003689 Homo sapi
	12	456.2	22.8	201460	8	AP003721	AP003721 Homo sapi
	13	451.2	22.6	195986	14	AC116933	AC116933 Papio anu
c	14	449.2	22.5	134210	8	AC005052	AC005052 Homo sapi
c	15	447.8	22.4	131943	8	AC005484	AC005484 Homo sapi
c	16	447.8	22.4	136119	8	HS1028D15	AL121886 Human DNA
	17	447.8	22.4	186197	14	AL627384	AL627384 Homo sapi
	18	446.2	22.3	163681	8	AL136992	AL136992 Human DNA
	19	445	22.3	117336	8	AC093171	AC093171 Homo sapi
	20	442.6	22.1	91927	8	AC004771	AC004771 Homo sapi
c	21	441.8	22.1	200430	8	AC011500	AC011500 Homo sapi
c	22	439.4	22.0	180076	14	AC022177	AC022177 Homo sapi
c	23	438.8	22.0	201227	14	AC160565	AC160565 Pan trogl
c	24	438.6	21.9	163704	14	AC141415	AC141415 Pan trogl
c	25	438.2	21.9	199045	14	AC149081	AC149081 Pan trogl
c	26	437.8	21.9	185574	14	AC161822	AC161822 Pan trogl
c	27	435.8	21.8	203790	8	AC010422	AC010422 Homo sapi
	28	435	21.8	119043	8	AC139451	AC139451 Homo sapi
c	29	435	21.8	163542	14	AC129071	AC129071 Pan trogl
c	30	434.8	21.8	96995	8	AC008149	AC008149 Homo sapi
	31	434.2	21.7	152659	8	AL591503	AL591503 Human DNA
c	32	433.8	21.7	199384	8	AP003531	AP003531 Homo sapi
	33	433.2	21.7	89448	8	AC005067	AC005067 Homo sapi
c	34	433.2	21.7	178976	14	AL391095	AL391095 Homo sapi
	35	432.6	21.6	156933	8	AC020750	AC020750 Homo sapi
c	36	432.2	21.6	81117	8	AC016898	AC016898 Homo sapi
c	37	432.2	21.6	192961	8	AP002770	AP002770 Homo sapi
c	38	432	21.6	153392	8	AL133548	AL133548 Human DNA
	39	432	21.6	235183	14	AC160951	AC160951 Colobus g
c	40	431.4	21.6	214031	14	AC160952	AC160952 Colobus g
c	41	431	21.6	134957	8	HSJ930L11	AL049759 Human DNA
c	42	430.8	21.6	44062	8	AC004517	AC004517 Homo sapi
c	43	430.8	21.6	118396	8	AC073138	AC073138 Homo sapi
	44	430.8	21.6	162701	14	AC073317	AC073317 Homo sapi
c	45	430.6	21.5	41613	8	AC004790	AC004790 Homo sapi

ALIGNMENTS

RESULT 1
AY438977
LOCUS AY438977 8368 bp DNA linear PRI 29-OCT-2003
DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
ACCESSION AY438977
VERSION AY438977.1 GI:37953284
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8368)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).
FEATURES Location/Qualifiers
source 1. .8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101. .257
/rpt_family="MIR"
/rpt_type=dispersed
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326. .538
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 354. .496

```
                /note="Region not scanned for variation"
gene            875. .6548
                /gene="PLA2G1B"
mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                /gene="PLA2G1B"
                /product="phospholipase A2, group IB (pancreas)"
CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                /gene="PLA2G1B"
                /codon_start=1
                /product="phospholipase A2, group IB (pancreas)"
                /protein_id="AAR05441.1"
                /db_xref="GI:37953285"
                /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
repeat_region   1174. .1307
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1286
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="g"
repeat_region   1319. .1588
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1343
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="a"
variation       1593
                /gene="PLA2G1B"
                /frequency="0.02"
                /replace="t"
repeat_region   1605. .1669
                /rpt_family="L2"
                /rpt_type=dispersed
misc_feature    1638. .2188
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
repeat_region   1670. .1975
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   1976. .2038
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2086. .2165
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2194. .2319
                /rpt_family="Alu"
                /rpt_type=dispersed
misc_feature    2319. .2453
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
variation       3027
                /gene="PLA2G1B"
                /frequency="0.25"
                /replace="t"
repeat_region   3165. .3431
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       3262
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3321
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
variation       3396
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="g"
variation       3630
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="c"
variation       3702
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="a"
variation       3768
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3818
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
repeat_region   3846. .3964
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   3977. .4280
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   4281. .4346
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   4347. .4634
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       4386
                /gene="PLA2G1B"
                /frequency="0.08"
                /replace="t"
repeat_region   4635. .4723
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5013
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="a"
repeat_region   5103. .5404
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5133
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5543
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5555
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       5556
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="g"
repeat_region   5577. .5629
                /rpt_family="L1"
                /rpt_type=dispersed
variation       5580
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
variation       5675
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
repeat_region   5779. .6083
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5929
```

```
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             100.0%;  Score 1998.6;  DB 8;  Length 8368;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1995;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       3574 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 3633

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              ||||||||||||||||||||||||||||||||||||||||||::||||||||||||||||||
Db       3634 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 3693

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3694 TCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 3753

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3754 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 3813

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3814 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 3873

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3874 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 3933

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3934 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 3993

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3994 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 4053

Qy        481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4054 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 4113

Qy        541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4114 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 4173

Qy        601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4174 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 4233

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4234 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 4293

Qy        721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4294 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 4353

Qy        781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4354 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 4413

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
              ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db       4414 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 4473

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4474 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 4533

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4534 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 4593

Qy       1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4594 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 4653

Qy       1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4654 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 4713

Qy       1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4714 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 4773

Qy       1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4774 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 4833

Qy       1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4834 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 4893

Qy       1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4894 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 4953

Qy       1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4954 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 5013

Qy       1441 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5014 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 5073

Qy       1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5074 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 5133

Qy       1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5134 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 5193

Qy       1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5194 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 5253

Qy       1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5254 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 5313

Qy       1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5314 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 5373

Qy       1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5374 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 5433

Qy       1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5434 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 5493

Qy       1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5494 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 5553

Qy       1981 ACKCTACATAAATATCTATTA 2001
              ||:||||||||||||||||||
Db       5554 ACTCTACATAAATATCTATTA 5574


RESULT 2
AX377239
LOCUS       AX377239               13612 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212562.
ACCESSION   AX377239
VERSION     AX377239.1  GI:19573528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 1 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .13612
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     variation       3845
                     /note="PS1: polymorphic base G or A"
     variation       3968
                     /note="PS2: polymorphic base A or G"
     variation       6060
                     /note="PS2: polymorphic base A or G"
     variation       6844
                     /note="PS4: polymorphic base G or A"
     variation       9531
                     /note="PS5: polymorphic base G or A"
ORIGIN

  Query Match             100.0%;  Score 1998.6;  DB 6;  Length 13612;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1996;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       6716 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 6775

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              ||||||||||||||||||||||||||||||||||||||||||::||||||||||||||||
Db       6776 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 6835

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6836 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 6895

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6896 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 6955

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6956 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 7015

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7016 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 7075

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7076 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 7135

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7136 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 7195

Qy        481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7196 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 7255

Qy        541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7256 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 7315

Qy        601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7316 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 7375

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7376 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 7435

Qy        721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7436 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 7495

Qy        781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7496 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 7555

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
              ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db       7556 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 7615

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 7675

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 7735

Qy       1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 7795

Qy       1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 7855

Qy       1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 7915

Qy       1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 7975

Qy       1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 8035

Qy       1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 8095

Qy       1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 8155

Qy       1441 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8156 CACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 8215

Qy        1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8216 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8275

Qy        1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8276 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8335

Qy        1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8336 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8395

Qy        1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8396 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8455

Qy        1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8456 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8515

Qy        1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8516 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 8575

Qy        1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8576 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 8635

Qy        1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8636 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 8695

Qy        1981 ACKCTACATAAATATCTATTA 2001
               ||:||||||||||||||||||
Db        8696 ACTCTACATAAATATCTATTA 8716
```

```
RESULT 3
AC003982/c
LOCUS       AC003982               122302 bp    DNA     linear   PRI 18-MAR-1999
DEFINITION  Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION   AC003982
VERSION     AC003982.1  GI:2769695
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122302)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 122302)
  AUTHORS   Bradshaw,H., Wu,X. and Ozersky,P.
  TITLE     The sequence of Homo sapiens PAC clone 166H1
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone was originally isolated in the laboratory of Professor
            Graeme Bell, Howard Hughes Medical Institute and Departments of
            Biochemistry and Molecular Biology, and Medicine, The University of
            Chicago, Chicago, IL, USA.  The clone was provided by the
            laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human
            Genetics, Oxford, UK.  Some contig information was also obtained
            from Yamagata et al., Nature 384:455-8 (1996).

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-1, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is 278C19; the clone sequenced to
            the right is 15E1.  Actual start of this clone is at base position
            1 of 166H1; actual end is at 122302 of 166H1.
FEATURES             Location/Qualifiers
     source          1..122302
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q"
                     /clone="166H1"
                     /clone_lib="RPCI-1"
     repeat_region   1..238
                     /rpt_family="Alu"
     repeat_region   239..265
                     /rpt_family="AT_rich"
     repeat_region   268..565
                     /rpt_family="Alu"
     repeat_region   1001..1309
                     /rpt_family="Alu"
     repeat_region   1440..1745
                     /rpt_family="Alu"
     repeat_region   1909..1940
                     /rpt_family="MER1_type"
     repeat_region   1955..1983
                     /rpt_family="AT_rich"
     repeat_region   2063..2362
                     /rpt_family="Alu"
     repeat_region   2516..2650
                     /rpt_family="Alu"
     repeat_region   2674..2730
                     /rpt_family="L1"
     repeat_region   2735..2788
                     /rpt_family="(GAAAA)n"
```

```
repeat_region   2790. .3077
                /rpt_family="Alu"
repeat_region   3081. .3139
                /rpt_family="L1"
repeat_region   3716. .4020
                /rpt_family="Alu"
repeat_region   4196. .4492
                /rpt_family="Alu"
gene            4560. .14034
                /gene="WUGSC:H_166H1.1"
CDS             join(4560. .5056,13418. .13748,13882. .14034)
                /gene="WUGSC:H_166H1.1"
                /note="unknown function; 60% similar to Z50177
                (PID:g927403) (PID:g927402); H_166H1.1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB95634.1"
                /db_xref="GI:2769696"
                /translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
                KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
                QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
                HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
                LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
                VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
repeat_region   5366. .5655
                /rpt_family="Alu"
repeat_region   5672. .5798
                /rpt_family="Alu"
repeat_region   5802. .5834
                /rpt_family="(TA)n"
repeat_region   5834. .5857
                /rpt_family="POLY_A"
repeat_region   5870. .6184
                /rpt_family="Alu"
repeat_region   6224. .6268
                /rpt_family="(CA)n"
repeat_region   6269. .6558
                /rpt_family="Alu"
repeat_region   6729. .7029
                /rpt_family="Alu"
repeat_region   7051. .7172
                /rpt_family="Alu"
repeat_region   7192. .7500
                /rpt_family="Alu"
repeat_region   7518. .7820
                /rpt_family="Alu"
repeat_region   8207. .8500
                /rpt_family="Alu"
repeat_region   8576. .8878
                /rpt_family="Alu"
repeat_region   9023. .9154
                /rpt_family="Alu"
repeat_region   9165. .9470
                /rpt_family="Alu"
repeat_region   9887. .10188
                /rpt_family="Alu"
repeat_region   10221. .10342
                /rpt_family="L1"
repeat_region   10408. .10707
                /rpt_family="Alu"
repeat_region   10713. .11014
                /rpt_family="Alu"
repeat_region   11020. .11110
                /rpt_family="Alu"
repeat_region   11146. .11386
                /rpt_family="Alu"
repeat_region   11208. .11386
                /rpt_family="Alu"
repeat_region   11391. .11515
                /rpt_family="Alu"
repeat_region   11568. .11705
                /rpt_family="Alu"
repeat_region   11737. .12032
                /rpt_family="Alu"
repeat_region   12034. .12066
                /rpt_family="7SLRNA"
repeat_region   12075. .12372
                /rpt_family="Alu"
repeat_region   12448. .12753
                /rpt_family="Alu"
repeat_region   12763. .12874
                /rpt_family="L2"
repeat_region   12999. .13303
                /rpt_family="Alu"
misc_feature    complement(13674. .13752)
                /note="match to EST N80020 (NID:g1242721) za91a08.s1"
misc_feature    complement(13840. .14195)
                /note="match to EST N80020 (NID:g1242721) za91a08.s1"
repeat_region   14196. .14229
                /rpt_family="AT_rich"
repeat_region   14240. .14339
                /rpt_family="U6"
repeat_region   14349. .14646
                /rpt_family="Alu"
repeat_region   14652. .14940
                /rpt_family="Alu"
```

Query Match 100.0%; Score 1998.6; DB 8; Length 122302;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1995; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db      26088 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 26029

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              ||||||||||||||||||||||||||||||||||||||||||::|||||||||||||||||
Db      26028 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 25969

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25968 TCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 25909

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25908 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 25849

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25848 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 25789

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25788 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 25729

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25728 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 25669

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25668 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 25609

Qy        481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25608 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 25549

Qy        541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25548 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 25489

Qy        601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25488 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 25429

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
```

```
Db      25428 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 25369

Qy        721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780

Db      25368 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 25309

Qy        781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840

Db      25308 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 25249

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900

Db      25248 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 25189

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960

Db      25188 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 25129

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020

Db      25128 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 25069

Qy       1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080

Db      25068 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 25009

Qy       1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140

Db      25008 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 24949

Qy       1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200

Db      24948 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 24889

Qy       1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260

Db      24888 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 24829

Qy       1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320

Db      24828 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 24769

Qy       1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380

Db      24768 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 24709

Qy       1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440

Db      24708 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 24649

Qy       1441 CACACACAGTTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500

Db      24648 CACACACAGTTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 24589

Qy       1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560

Db      24588 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 24529

Qy       1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620

Db      24528 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 24469

Qy       1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680

Db      24468 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 24409

Qy       1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740

Db      24408 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 24349

Qy       1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800

Db      24348 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 24289

Qy       1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860

Db      24288 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 24229

Qy       1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920

Db      24228 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 24169

Qy       1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980

Db      24168 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 24109

Qy       1981 ACKCTACATAAATATCTATTA 2001

Db      24108 ACTCTACATAAATATCTATTA 24088
```

```
RESULT 4
AC078926/c
LOCUS       AC078926               220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match             100.0%;  Score 1998.6;  DB 14;  Length 220384;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1995;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db      71311 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 71252

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||::||||||||||||
Db      71251 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 71192

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71191 TCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 71132

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71131 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 71072

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71071 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 71012

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71011 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 70952

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70951 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTTGTTGTTGT 70892

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70891 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 70832

Qy        481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70831 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 70772

Qy        541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70771 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 70712

Qy        601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70711 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 70652

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70651 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 70592
```

```
Qy       721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70591 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 70532

Qy       781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70531 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 70472

Qy       841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
            |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db     70471 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 70412

Qy       901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70411 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 70352

Qy       961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70351 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 70292

Qy      1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70291 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 70232

Qy      1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70231 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 70172

Qy      1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70171 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 70112

Qy      1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70111 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 70052

Qy      1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70051 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 69992

Qy      1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69991 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 69932

Qy      1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69931 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 69872

Qy      1441 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69871 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 69812

Qy      1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69811 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 69752

Qy      1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69751 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 69692

Qy      1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69691 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 69632

Qy      1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69631 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 69572

Qy      1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69571 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 69512
```

```
Qy      1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69511 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 69452

Qy      1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69451 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 69392

Qy      1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69391 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 69332

Qy      1981 ACKCTACATAAATATCTATTA 2001
            ||:||||||||||||||||||
Db     69331 ACTCTACATAAATATCTATTA 69311
```

RESULT 5
AC073930
LOCUS AC073930 189729 bp DNA linear HTG 01-MAR-2002
DEFINITION Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC073930
VERSION AC073930.23 GI:18958593
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189729)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished

```
REFERENCE   2  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14861669.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBWS
                Center clone name: RP11-144B2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191640 bases at least Q40
                Consensus quality: 195135 bases at least Q30
                Consensus quality: 197786 bases at least Q20
                Estimated insert size: 193340; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    64773: contig of 64773 bp in length
            *   64774    64873: gap of unknown length
            *   64874    99097: contig of 34224 bp in length
            *   99098    99197: gap of unknown length
            *   99198   135331: contig of 36134 bp in length
            *  135332   135431: gap of unknown length
            *  135432   161344: contig of 25913 bp in length
            *  161345   161444: gap of unknown length
            *  161445   171231: contig of 9787 bp in length
            *  171232   171331: gap of unknown length
            *  171332   180922: contig of 9591 bp in length
            *  180923   181022: gap of unknown length
            *  181023   187271: contig of 6249 bp in length
            *  187272   187371: gap of unknown length
            *  187372   189729: contig of 2358 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-144B2"
     gap             64774. .64873
                     /estimated_length=unknown
     gap             99098. .99197
                     /estimated_length=unknown
     gap             135332. .135431
                     /estimated_length=unknown
     gap             161345. .161444
                     /estimated_length=unknown
     gap             171232. .171331
                     /estimated_length=unknown
     gap             180923. .181022
                     /estimated_length=unknown
     gap             187272. .187371
                     /estimated_length=unknown
ORIGIN

  Query Match           90.6%;  Score 1811.2;  DB 14;  Length 189729;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1810;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        190 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171332 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 171391

Qy        250 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171392 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 171451

Qy        310 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171452 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 171511

Qy        370 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 429
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171512 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 171571

Qy        430 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171572 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 171631

Qy        490 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 549
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171632 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 171691

Qy        550 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171692 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 171751

Qy        610 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171752 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 171811

Qy        670 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171812 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 171871

Qy        730 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171872 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 171931

Qy        790 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171932 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 171991

Qy        850 TCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 909
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db     171992 TCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 172051

Qy        910 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172052 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 172111

Qy        970 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172112 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 172171

Qy       1030 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 1089
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172172 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 172231

Qy       1090 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 1149
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172232 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 172291

Qy       1150 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 1209
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172292 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 172351

Qy       1210 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 1269
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172352 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 172411

Qy       1270 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 1329
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172412 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 172471

Qy       1330 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 1389
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172472 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 172531

Qy       1390 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 1449
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172532 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 172591

Qy       1450 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1509
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172592 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 172651

Qy       1510 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1569
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172652 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 172711

Qy       1570 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1629
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172712 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 172771

Qy       1630 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1689
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172772 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 172831

Qy       1690 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1749
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172832 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 172891

Qy       1750 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 1809
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172892 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 172951

Qy       1810 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 1869
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172952 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 173011

Qy       1870 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 1929
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173012 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 173071

Qy       1930 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACKCTACAT 1989
                ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db     173072 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACGCTACAT 173131

Qy       1990 AAATATCTATTA 2001
                ||||||||||||
Db     173132 AAATATCTATTA 173143


RESULT 6
AC068981/c
LOCUS       AC068981               178887 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-580I7, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC068981
VERSION     AC068981.3  GI:9958291
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:8748938.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0580I07
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157045 bases at least Q40
            Consensus quality: 163476 bases at least Q30
            Consensus quality: 167183 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 174828; sum-of-contigs
            Quality coverage: 3.01 in Q20 bases; agarose-fp
            Quality coverage: 3.40 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1242: contig of 1242 bp in length
            *     1243     1342: gap of unknown length
            *     1343     2435: contig of 1093 bp in length
            *     2436     2535: gap of unknown length
            *     2536     3664: contig of 1129 bp in length
            *     3665     3764: gap of unknown length
            *     3765     4864: contig of 1100 bp in length
            *     4865     4964: gap of unknown length
            *     4965     7695: contig of 2731 bp in length
            *     7696     7795: gap of unknown length
            *     7796    10187: contig of 2392 bp in length
            *    10188    10287: gap of unknown length
            *    10288    12303: contig of 2016 bp in length
            *    12304    12403: gap of unknown length
            *    12404    14390: contig of 1987 bp in length
            *    14391    14490: gap of unknown length
            *    14491    16285: contig of 1795 bp in length
            *    16286    16385: gap of unknown length
            *    16386    18364: contig of 1979 bp in length
            *    18365    18464: gap of unknown length
            *    18465    20794: contig of 2330 bp in length
            *    20795    20894: gap of unknown length
            *    20895    23473: contig of 2579 bp in length
            *    23474    23573: gap of unknown length
            *    23574    25387: contig of 1814 bp in length
            *    25388    25487: gap of unknown length
            *    25488    28870: contig of 3383 bp in length
            *    28871    28970: gap of unknown length
            *    28971    32203: contig of 3233 bp in length
```

```
          *    32204    32303: gap of unknown length
          *    32304    32762: contig of 459 bp in length
          *    32763    32862: gap of unknown length
          *    32863    36428: contig of 3566 bp in length
          *    36429    36528: gap of unknown length
          *    36529    39558: contig of 3030 bp in length
          *    39559    39658: gap of unknown length
          *    39659    44249: contig of 4591 bp in length
          *    44250    44349: gap of unknown length
          *    44350    48970: contig of 4621 bp in length
          *    48971    49070: gap of unknown length
          *    49071    54728: contig of 5658 bp in length
          *    54729    54828: gap of unknown length
          *    54829    60549: contig of 5721 bp in length
          *    60550    60649: gap of unknown length
          *    60650    66747: contig of 6098 bp in length
          *    66748    66847: gap of unknown length
          *    66848    73820: contig of 6973 bp in length
          *    73821    73920: gap of unknown length
          *    73921    80722: contig of 6802 bp in length
          *    80723    80822: gap of unknown length
          *    80823    86541: contig of 5719 bp in length
          *    86542    86641: gap of unknown length
          *    86642    91906: contig of 5265 bp in length
          *    91907    92006: gap of unknown length
          *    92007    99976: contig of 7970 bp in length
          *    99977   100076: gap of unknown length
          *   100077   108154: contig of 8078 bp in length
          *   108155   108254: gap of unknown length
          *   108255   114943: contig of 6689 bp in length
          *   114944   115043: gap of unknown length
          *   115044   122680: contig of 7637 bp in length
          *   122681   122780: gap of unknown length
          *   122781   130297: contig of 7517 bp in length
          *   130298   130397: gap of unknown length
          *   130398   138520: contig of 8123 bp in length
          *   138521   138620: gap of unknown length
          *   138621   146448: contig of 7828 bp in length
          *   146449   146548: gap of unknown length
          *   146549   152895: contig of 6347 bp in length
          *   152896   152995: gap of unknown length
          *   152996   162254: contig of 9259 bp in length
          *   162255   162354: gap of unknown length
          *   162355   178887: contig of 16533 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178887
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-580I7"
     misc_feature    1. .1242
                     /note="assembly_name:Contig22"
     gap             1243. .1342
                     /estimated_length=unknown
     misc_feature    1343. .2435
                     /note="assembly_name:Contig24"
     gap             2436. .2535
                     /estimated_length=unknown
     misc_feature    2536. .3664
                     /note="assembly_name:Contig25"
     gap             3665. .3764
                     /estimated_length=unknown
     misc_feature    3765. .4864
                     /note="assembly_name:Contig27"
     gap             4865. .4964
                     /estimated_length=unknown
     misc_feature    4965. .7695
                     /note="assembly_name:Contig28"
     gap             7696. .7795
                     /estimated_length=unknown
     misc_feature    7796. .10187
                     /note="assembly_name:Contig30"
     gap             10188. .10287
                     /estimated_length=unknown
     misc_feature    10288. .12303
                     /note="assembly_name:Contig31"
     gap             12304. .12403
                     /estimated_length=unknown
     misc_feature    12404. .14390
                     /note="assembly_name:Contig32"
     gap             14391. .14490
                     /estimated_length=unknown
     misc_feature    14491. .16285
                     /note="assembly_name:Contig33"
     gap             16286. .16385
                     /estimated_length=unknown
     misc_feature    16386. .18364
                     /note="assembly_name:Contig34"
     gap             18365. .18464
                     /estimated_length=unknown
     misc_feature    18465. .20794
                     /note="assembly_name:Contig35"
     gap             20795. .20894
                     /estimated_length=unknown
     misc_feature    20895. .23473
                     /note="assembly_name:Contig36"
     gap             23474. .23573
                     /estimated_length=unknown
     misc_feature    23574. .25387
                     /note="assembly_name:Contig37"
     gap             25388. .25487
                     /estimated_length=unknown
     misc_feature    25488. .28870
                     /note="assembly_name:Contig38"
     gap             28871. .28970
                     /estimated_length=unknown
     misc_feature    28971. .32203
                     /note="assembly_name:Contig39
                     clone_end:SP6
                     vector_side:right"
     gap             32204. .32303
                     /estimated_length=unknown
     misc_feature    32304. .32762
                     /note="assembly_name:Contig12"
     gap             32763. .32862
                     /estimated_length=unknown
     misc_feature    32863. .36428
                     /note="assembly_name:Contig40"
     gap             36429. .36528

  Query Match             23.1%;  Score 462.6;  DB 14;  Length 178887;
  Best Local Similarity   62.1%;  Pred. No. 2.3e-85;
  Matches  890;  Conservative    1;  Mismatches  485;  Indels   57;  Gaps    8;

Qy        400 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 459
              || || |||| |||||||||||      | | ||  |||||||| |||||||  |||
Db      90620 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 90561

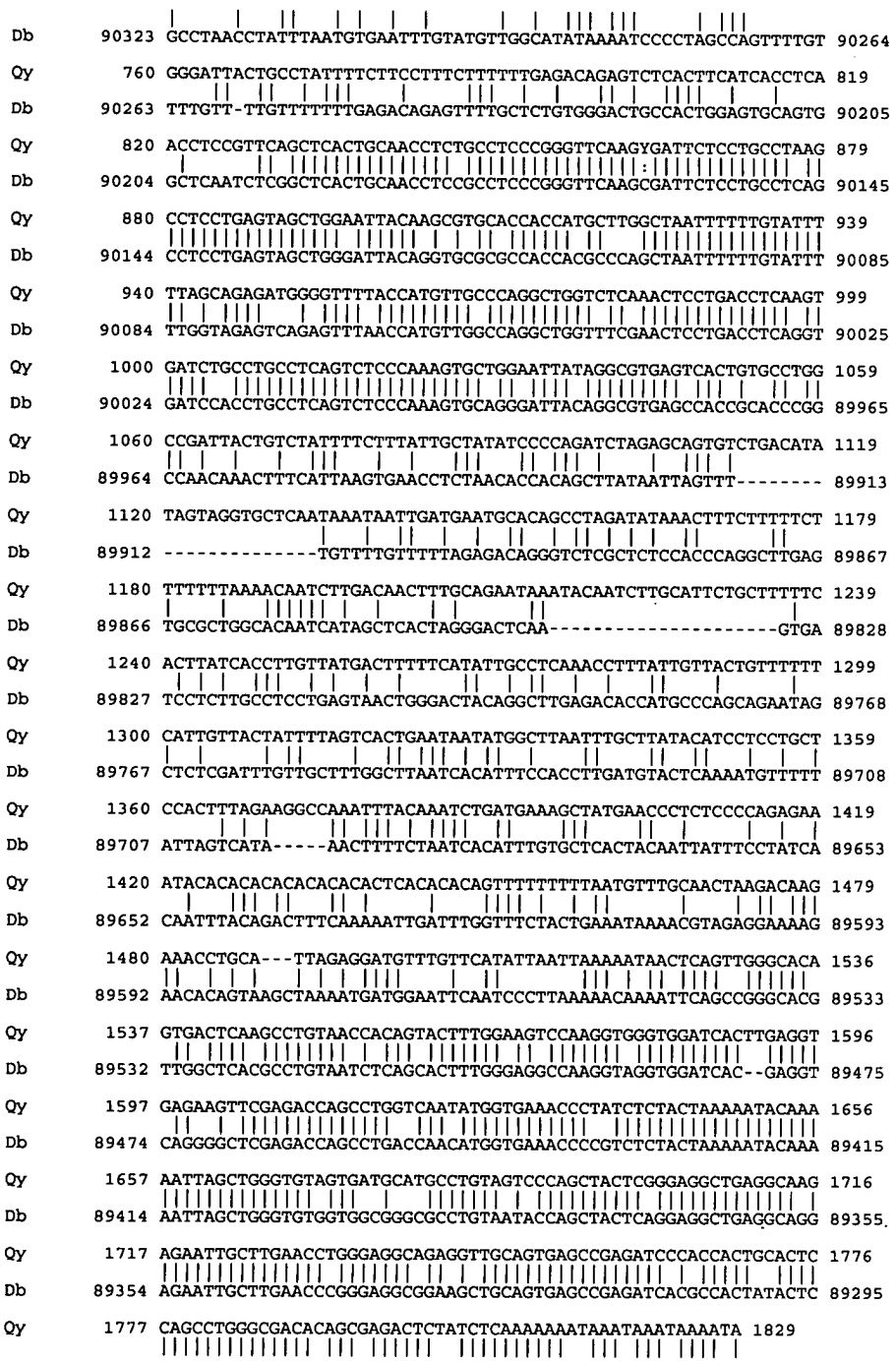
Qy        460 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 519
              ||||||  | | |  | ||||||||  |||  ||||||| |||||||  |||  || ||
Db      90560 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 90501

Qy        520 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 579
              |||||| ||||||||||||||||||  | |  || ||||| || | |||||||
Db      90500 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 90441

Qy        580 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 639
              |  |||| ||||||||| |||| || | | |||| ||| |||  | | |||||||||
Db      90440 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 90383

Qy        640 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 699
              |||| | |||| |||  ||  || |||||||||||| |||||||||||| ||||||||||
Db      90382 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 90324

Qy        700 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 759
```



Db 89294 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAACA 89242

RESULT 7
AC009244
LOCUS AC009244 190162 bp DNA linear PRI 26-MAY-2002
DEFINITION Homo sapiens chromosome 7 clone RP11-305M3, complete sequence.
ACCESSION AC009244
VERSION AC009244.25 GI:21217399
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 190162)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190162)
AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 190162)
AUTHORS Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 4 (bases 1 to 190162)
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On May 26, 2002 this sequence version replaced gi:11038533.
---------- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
---------- Project Information
Center project name: chr-7
Center clone name: RP11-305M3 (djs195)
---------- Summary Statistics
Assembly program: Phrap; version 0.990319
Consensus quality: 189969 bases at least Q40
Consensus quality: 190110 bases at least Q30
Consensus quality: 190161 bases at least Q20
Insert size: 190162; sum-of-contigs
Quality coverage: 9.7x in Q20 bases; sum-of-contigs
----------.
Overlapping Sequences:
5': RP11-382M23 (UWGC:djs734g) AC093149 1340-bp overlap
3': RP11-448A19 (UWGC:djs705) AC078846 8765-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred

quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
21251	20929	12417	12174	8696	8794
12148	11918	2067	2041	6	<800
199	<800	5666	5826	1775	1773
3275	3221	254	<800	2000	1991
20379	20929	3818	3836	6313	6324
15555	15150	747	735	3256	3226
675	<800	9196	9237	325	<800
5493	5475	4154	4101	12989	12723
516	<800	5820	5826	5955	5951
5810	5831	118	<800	611	<800
9488	9435	2524	2359	1995	1991
2908	2935	8844	8768	11945	12009
779	840	332	<800	5577	5593
821	840	564	<800	4139	4124
315	<800	26	<800	955	940
3640	3671	4675	4644	2961	2940
428	<800	1801	1803	5264	5205
1144	1119	117	<800	1535	1534
4684	4717	6797	6799	3313	3383
20	<800	3146	3177	3760	3761
492	<800	2376	2359	3061	3110
10935	10879	210	<800	9605	9505
5921	6010	5135	5110	335	<800
529	<800	744	735	2190	2194
4354	4384	881	947	5117	5205
1311	1293	413	<800	13558	13434
10	<800	6389	6456	9367	9505
374	<800	296	<800	8080	8019
1497	1480	7160	7119	4276	4243
4079	4083	4846	4822	7830	8019
337	<800	2914	2924	231	<800
4483	4384	3803	3836	533	<800
1530	1480	1832	1803	668	<800
249	<800	674	<800	12314	12009
8831	8995	2515	2513	18561	18820
244	<800	2163	2137	11006	11013
1019	1000	466	<800	8756	8794
841	840	927	947	7	<800
8114	8086	5963	5826		
9184	9435	7440	7660		
7121	7263	5159	4822		
1122	1119	677	<800		
49	<800	1399	1373		
11928	11918	1565	1554		
1499	1480	2388	2513		
2585	2575	4826	4822		
699	<800	7661	7660		
		3624	3705		
		8693	8768		
		1749	1803		
		1642	1635		
		5295	5362		
		10628	10715		
		3227	3177		
		589	<800		

Query Match 23.1%; Score 462.6; DB 8; Length 190162;
Best Local Similarity 61.3%; Pred. No. 2.3e-85;
Matches 900; Conservative 1; Mismatches 520; Indels 46; Gaps 8;

```
Qy      404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
            ||||  ||| ||      |||| || | |||||||| |||||||||| | |  ||||
Db    31687 TTTTCCTTTTTTTAAAAGTAGAAACGGAGTCTTGCTGGGTCACCCAGGCTGGGGTGCAGT 31746

Qy      464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
            |||||  | |   |||||||| |||  |  ||||  |  || | ||||  |  || |||
Db    31747 AGTGCAGTCGTGTTTCACTGCATCCTTCAATTCCTAGGAACAGGTGATCCTCCCACTTCA 31806

Qy      524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
```

```
Db      31807 GTCTTCTGAGTAGCTGGGACTTCAGGCATGTGCCACCATGCCCAGCTAATTTAAAAAAAA 31866

Qy        584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643

Db      31867 TGTTTGAAGACACAGGGTCTTGCTGTGTTGCTCAGGCTGGTCTCAAACAGCCAGCCTCAA 31926

Qy        644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703

Db      31927 GTGATCCTCCTGCCTCATCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCC 31986

Qy        704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 763

Db      31987 AGCCCATTTCTTTTTTCTGTTTTTTTTTTTAAATTTAATTTAATTTTTATTTTTTATTTT 32046

Qy        764 TTACTGCCTATTTTCTTCCTT-TCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 822

Db      32047 TATTTTTATTTTTATTTATTTATTTTTTTTGAGATGAAGTCTCACTCTGTCGCCCAGGCT 32106

Qy        823 TCCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 868

Db      32107 GGAGTGCAGTGGCATGATCTCGGCTCACTGAAACCTCTGCCTCCCAGGTTCAAGCGATTC 32166

Qy        869 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAAT 928

Db      32167 TCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCCTCCACTATACCCGGCTAA- 32225

Qy        929 TTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCC 988

Db      32226 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGTCAGGCTGGTCTCGAACTCC 32285

Qy        989 TGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTC 1048

Db      32286 TGACCTCAGATGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCC 32345

Qy       1049 ACTGTGCCTGGCCGATTACTGTCTATTTT---CTTTATTGCTATATCCCCAGATCTAGAG 1105

Db      32346 ACCGCACCCGGCTGCCCCATTTTTTTTTTTAAACAAATGCTTACTGACCTACGCTAAGAG 32405

Qy       1106 CAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATA 1165

Db      32406 CTGTAAGGGCAAATGAGACAATGTCCTTACCCTCAGAATCTTATATTTTGGTGAGATAAT 32465

Qy       1166 AACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTG 1225

Db      32466 ACACATATATAAATAATGATGATACAAGGACATATATAATAAATGCCATGACAGATTATA 32525

Qy       1226 CATTCTGCTTTTTCAC-TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1284

Db      32526 AATCCATATTCACTACACAAACAAGAGGCTTAGGAAAGTTCAGATTCCCACTTTGGGAAG 32585

Qy       1285 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1344

Db      32586 CTGAGGCAGGTGGATCACCTGAGGTCAGGAATTCA--AGACCAGCGTGGCCAACATGGTG 32643

Qy       1345 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1404

Db      32644 AAACCCTGTCTTTACTAAGAGTACAAAAATTAGCTGAGTGTGGTGGTGCAGGCCTGTAAT 32703

Qy       1405 CCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTT 1464

Db      32704 CCTAGCTACTCATGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGCTGCA 32763

Qy       1465 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTG---------------------- 1502

Db      32764 GTGAGCCGAGATCGCGCCACTGCACTCTAGCCTGGGTGACAGAGTGAGACTCCATCTCAA 32823

Qy       1503 TTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTAC 1562

Db      32824 AAAAAAAAAAAAAAAAAAAAAGAGGCCAGGCGCGGTGGCTCACACCTGTAATCCCAGCAC 32883

Qy       1563 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAA 1622

Db      32884 TTTGGGAGCCTGAGGTGGGCGGATCA--TGAGGTCAGAAGTTCGAGACCAGCCTGGCCAA 32941

Qy       1623 TATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGC 1682

Db      32942 TATGGTGAAACCCCGTCTTTACTAAAAGTACAAAAATTAGCTGGGCGTGGTGCCACGCAC 33001

Qy       1683 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 1742

Db      33002 CTGTAGTCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGA 33061

Qy       1743 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCT 1802

Db      33062 GGTTGCAGTGAGCCGAGATTATGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCC 33121

Qy       1803 ATCTCAAAAAAATAAATAAATAAAATA 1829

Db      33122 GTCTCAAAAAAAAAAAAGAAAAGAAAA 33148
```

```
RESULT 8
AP000589
LOCUS       AP000589              123192 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-105N5 map 11q13, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP000589
VERSION     AP000589.3  GI:8118795
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 123,192 genomic DNA of 11q13
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997479.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-105N5
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 115116 bases at least Q40
                Consensus quality: 118340 bases at least Q30
                Consensus quality: 120191 bases at least Q20
                Insert size: 121392; sum-of-contigs
                Quality coverage: 5.08x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
```

```
            1     15827 contig of  15827 bp in length
        15928     30484 contig of  14557 bp in length
        30585     40602 contig of  10018 bp in length
        40703     51912 contig of  11210 bp in length
        52013     60022 contig of   8010 bp in length
        60123     70438 contig of  10316 bp in length
        70539     77945 contig of   7407 bp in length
        78046     85717 contig of   7672 bp in length
        85818     92399 contig of   6582 bp in length
        92500     98215 contig of   5716 bp in length
        98316    104148 contig of   5833 bp in length
       104249    109166 contig of   4918 bp in length
       109267    112457 contig of   3191 bp in length
       112558    115076 contig of   2519 bp in length
       115177    117063 contig of   1887 bp in length
       117164    118652 contig of   1489 bp in length
       118753    120425 contig of   1673 bp in length
       120526    121843 contig of   1318 bp in length
       121944    123192 contig of   1249 bp in length
Sequence updated (18-Oct-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     15827: contig of 15827 bp in length
*     15828     15927: gap of 100 bp
*     15928     30484: contig of 14557 bp in length
*     30485     30584: gap of 100 bp
*     30585     40602: contig of 10018 bp in length
*     40603     40702: gap of 100 bp
*     40703     51912: contig of 11210 bp in length
*     51913     52012: gap of 100 bp
*     52013     60022: contig of 8010 bp in length
*     60023     60122: gap of 100 bp
*     60123     70438: contig of 10316 bp in length
*     70439     70538: gap of 100 bp
*     70539     77945: contig of 7407 bp in length
*     77946     78045: gap of 100 bp
*     78046     85717: contig of 7672 bp in length
*     85718     85817: gap of 100 bp
*     85818     92399: contig of 6582 bp in length
*     92400     92499: gap of 100 bp
*     92500     98215: contig of 5716 bp in length
*     98216     98315: gap of 100 bp
*     98316    104148: contig of 5833 bp in length
*    104149    104248: gap of 100 bp
*    104249    109166: contig of 4918 bp in length
*    109167    109266: gap of 100 bp
*    109267    112457: contig of 3191 bp in length
*    112458    112557: gap of 100 bp
*    112558    115076: contig of 2519 bp in length
*    115077    115176: gap of 100 bp
*    115177    117063: contig of 1887 bp in length
*    117064    117163: gap of 100 bp
*    117164    118652: contig of 1489 bp in length
*    118653    118752: gap of 100 bp
*    118753    120425: contig of 1673 bp in length
*    120426    120525: gap of 100 bp
*    120526    121843: contig of 1318 bp in length
*    121844    121943: gap of 100 bp
*    121944    123192: contig of 1249 bp in length.
FEATURES             Location/Qualifiers
     source          1. .123192
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q13"
                     /clone="CMB9-105N5"
     misc_feature    1. .15827
                     /note="assembly_fragment"
     misc_feature    15928. .30484
                     /note="assembly_fragment"
     misc_feature    30585. .40602
                     /note="assembly_fragment"
     misc_feature    40703. .51912
                     /note="assembly_fragment"
     misc_feature    52013. .60022
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    60123. .70438
                     /note="assembly_fragment"
     misc_feature    70539. .77945
                     /note="assembly_fragment"
     misc_feature    78046. .85717
                     /note="assembly_fragment"
     misc_feature    85818. .92399
                     /note="assembly_fragment"
     misc_feature    92500. .98215
                     /note="assembly_fragment"
     misc_feature    98316. .104148
                     /note="assembly_fragment"
     misc_feature    104249. .109166
                     /note="assembly_fragment"
     misc_feature    109267. .112457
                     /note="assembly_fragment"
     misc_feature    112558. .115076
                     /note="assembly_fragment"
     misc_feature    115177. .117063
                     /note="assembly_fragment"
     misc_feature    117164. .118652
                     /note="assembly_fragment"
     misc_feature    118753. .120425
                     /note="assembly_fragment"
     misc_feature    120526. .121843
                     /note="assembly_fragment"
     misc_feature    121944. .123192
                     /note="assembly_fragment"
ORIGIN

  Query Match           23.0%;  Score 460.6;  DB 14;  Length 123192;
  Best Local Similarity   59.8%;  Pred. No. 6.3e-85;
  Matches  882;  Conservative    1;  Mismatches  575;  Indels   17;  Gaps    6;

Qy        398 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 456
              |||    |||||||||| || || ||| ||||||| || |  ||| |||  ||||||| ||
Db      63120 TTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 63179

Qy        457 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 516
              |  || |||  ||| |   | ||||  || ||||  ||||| | | ||||||| |   |
Db      63180 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 63239

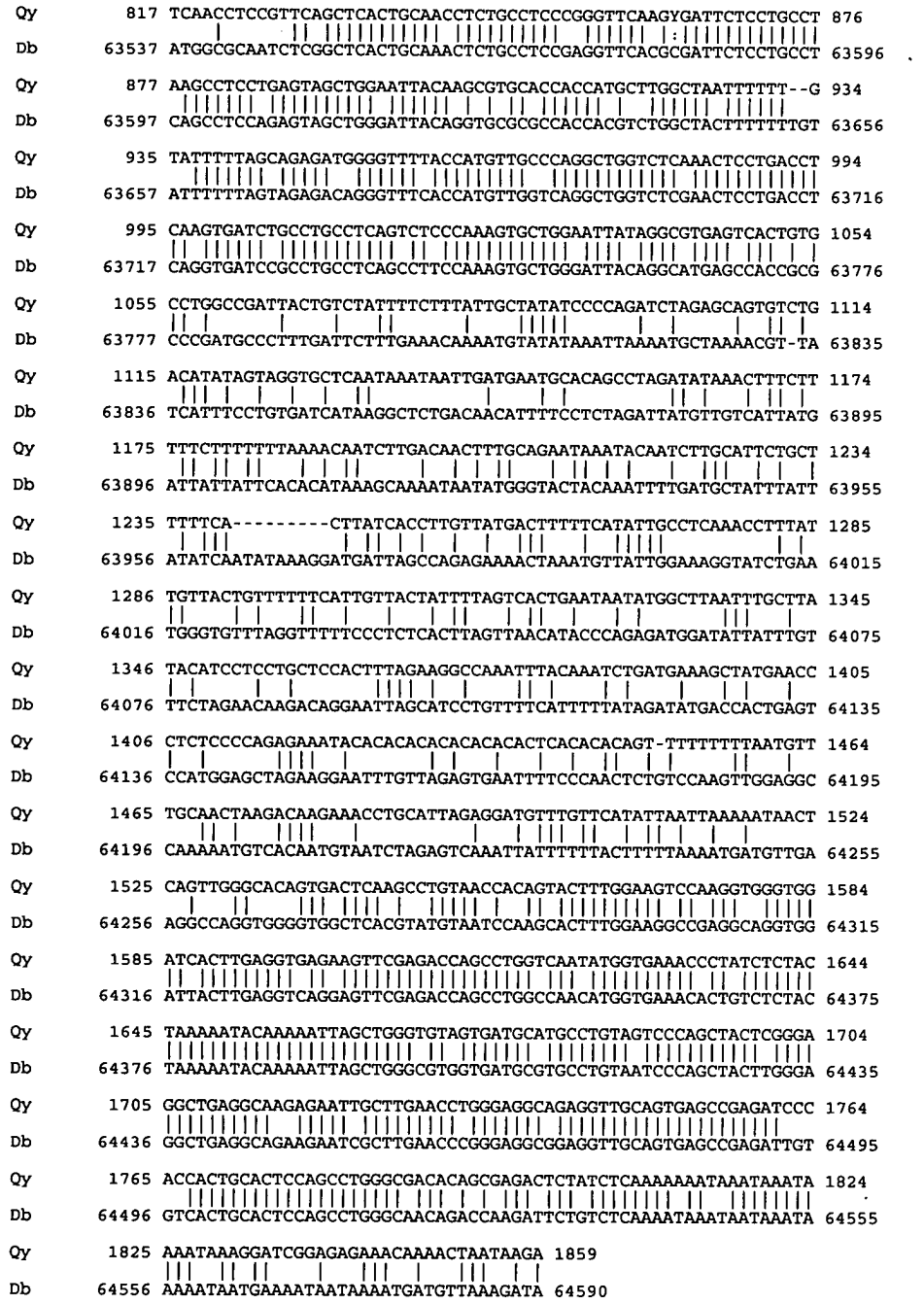
Qy        517 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 576
              |||||||||||| |||||||| || |||  || ||| ||||||||||  ||| |||
Db      63240 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 63299

Qy        577 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 636
              |   || || |||||||| |||| |  | | ||||| |||||| |||| | ||||||
Db      63300 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 63356

Qy        637 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 696
              | ||| |  |||| || |||| ||| ||||||||||||||||||| ||| ||||||||
Db      63357 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 63416

Qy        697 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 756
              || |  |||| |  || |  |||   |      ||   ||    || |           |
Db      63417 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 63476

Qy        757 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 816
              |      ||     || |||| ||        | ||           |  ||  |  |
Db      63477 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 63536
```



```
RESULT 9
AC022408
LOCUS       AC022408              150266 bp    DNA     linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-118J16 map 15q15, complete
            sequence.
ACCESSION   AC022408
VERSION     AC022408.6  GI:15217198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Aug 21, 2001 this sequence version replaced gi:11225361.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Note:data from AC021753 [Drafting center UWMSC] and AC012652
            [Drafting center UWMSC] and 580I7 [Drafting center WUGSC] were
            added for finishing.
FEATURES             Location/Qualifiers
     source          1. .150266
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-118J16"
                     /clone_lib="RPCI human BAC library 11"
                     /note="This clone overlaps RP11-46M12 AC012652 and
                     RP11-129I12 AC021753 and RP11-580I7 AC068981.Data from
                     overlapping BACs were combined and the consensus sequence
                     determined from RP11-118J16."
     misc_feature    1. .150266
                     /note="overlap with RP11-580I7 AC068981."
     misc_feature    1. .100397
                     /note="overlap with RP11-129I12 AC021753."
     unsure          49355. .49359
                     /note="low quality data."
     misc_feature    100628. .150266
                     /note="overlap with RP11-46M12 AC012652."
ORIGIN

  Query Match             23.0%;  Score 460.4;  DB 8;  Length 150266;
```

```
 Best Local Similarity   62.0%;  Pred. No. 6.8e-85;
 Matches  889;  Conservative    1;  Mismatches  487;  Indels   57;  Gaps    8;

Qy       400 TGGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 459
Db     52138 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 52197

Qy       460 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 519
Db     52198 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 52257

Qy       520 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 579
Db     52258 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 52317

Qy       580 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 639
Db     52318 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 52375

Qy       640 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 699
Db     52376 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCAA 52434

Qy       700 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 759
Db     52435 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 52494

Qy       760 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 819
Db     52495 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 52553

Qy       820 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 879
Db     52554 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 52613

Qy       880 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 939
Db     52614 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 52673

Qy       940 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 999
Db     52674 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 52733

Qy      1000 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1059
Db     52734 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 52793

Qy      1060 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1119
Db     52794 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 52845

Qy      1120 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1179
Db     52846 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 52891

Qy      1180 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1239
Db     52892 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 52930

Qy      1240 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1299
Db     52931 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 52990

Qy      1300 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1359
Db     52991 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 53050

Qy      1360 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1419
Db     53051 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 53105

Qy      1420 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1479
Db     53106 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 53165

Qy      1480 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1536
Db     53166 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 53225

Qy      1537 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1596
Db     53226 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 53283

Qy      1597 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1656
Db     53284 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 53343

Qy      1657 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1716
Db     53344 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 53403

Qy      1717 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1776
Db     53404 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACACCACTATACTC 53463

Qy      1777 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 1830
Db     53464 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAACATTCA 53517
```

```
RESULT 10
AC021753
LOCUS       AC021753               167996 bp    DNA     linear   PRI 29-OCT-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-129I12 map 15q15, complete
            sequence.
ACCESSION   AC021753
VERSION     AC021753.7  GI:16507213
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 167996)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167996)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 167996)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Oct 29, 2001 this sequence version replaced gi:11136698.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
```

```
            Note: data from AC022408 [Drafting center UWMSC] and AC020661
          were added for fimishing.
FEATURES             Location/Qualifiers
     source          1. .167996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-129I12"
                     /clone_lib="RPCI Human BAC library 11"
                     /note="This clone overlaps RP11-118J16 AC022408 and
                     RP11-540O11 AC020661. Data from overlapping BACs were
                     combined and the consensus sequence determined from
                     RP11-129I12 to the extent possible."
     misc_feature    1. .63402
                     /note="overlap with RP11-540O11 AC020661."
     misc_feature    67596. .167996
                     /note="overlap with RP11-118J16 AC022408."
ORIGIN

  Query Match             23.0%;  Score 459.4;  DB 8;  Length 167996;
  Best Local Similarity   62.0%;  Pred. No. 1.1e-84;
  Matches  888;  Conservative    1;  Mismatches  487;  Indels   57;  Gaps    8;

Qy        400 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 459
              || || |||| |||||||||||       | | ||  |||||||| |||||||  |||
Db     119755 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 119814

Qy        460 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 519
              ||||||  | | |   | ||||||||  |||  |||||| |||||||  |||  || ||
Db     119815 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 119874

Qy        520 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 579
              |||||| |||||||||||||||||||  | |  || ||||| || | |||||||
Db     119875 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 119934

Qy        580 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 639
              |  |||| |||||||| |||| || | | |||| ||| |||  | | |||||||||
Db     119935 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 119992

Qy        640 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 699
              |||| | |||| |||  ||  || |||||||||| |||||||||| ||||||||||
Db     119993 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 120051

Qy        700 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 759
              |      |   ||   | | |       | |  ||| |||      | |||
Db     120052 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 120111

Qy        760 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 819
                ||  ||  | |||    |     |||  |       |     ||||  |   |
Db     120112 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 120170

Qy        820 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 879
              |     || |||||||||||||| ||||||||||||||||:|||||||||||||| ||
Db     120171 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 120230

Qy        880 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 939
              |||||||||||||||| |||||| | | || |||||| ||   ||||||||||||||||
Db     120231 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 120290

Qy        940 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 999
              || | ||||   | |||| ||||||||| |||||||||| || ||||||||||||| ||
Db     120291 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 120350

Qy       1000 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1059
              ||||  |||||||||||||||||||||| || |||| ||||||||| ||| |  || ||
Db     120351 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 120410

Qy       1060 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1119
              || |  |   | |||   |  |   |||   || ||| |     |  ||| |
Db     120411 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 120462

Qy       1120 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1179
                                |  |  ||  |  |  ||  | || | | |  ||    ||
Db     120463 ----------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 120508

Qy       1180 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1239
              |   |  |||||| |  |    | |      ||                       |
Db     120509 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 120547

Qy       1240 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1299
              | | |  ||| | |  |  |    ||  | ||  |  |   ||    |     |
Db     120548 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 120607

Qy       1300 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1359
              |  |      | ||    |  || ||| | ||  |    ||    ||  ||   |  |
Db     120608 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 120667

Qy       1360 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1419
              | | |    || ||| | ||||  ||    |||    ||  |    |  |  |
Db     120668 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 120722

Qy       1420 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1479
              |   ||| ||  || |    |  ||||  |  | ||     |   | || |||
Db     120723 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 120782

Qy       1480 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1536
              ||  | | |   | | | |||||    |  ||       ||| | || ||||  ||||||
Db     120783 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 120842

Qy       1537 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1596
              ||| |||| |||||||| | ||| ||||||| || ||||||| ||||||||||  |||||
Db     120843 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 120900

Qy       1597 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1656
              ||  | |||||||||||||  ||| ||||||||||||  ||||||||||||||||||
Db     120901 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 120960

Qy       1657 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1716
              |||||||||||||||  |||  |  ||||||| | ||||||||| ||||||||||||| |
Db     120961 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 121020

Qy       1717 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1776
              |||||||||||||| ||||||| || | |||||||||||||||| | |||||  ||||
Db     121021 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACGCCACTATACTC 121080

Qy       1777 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 1829
              |||||||||||||| ||| ||||||  |||||||||  ||| ||| |||| |
Db     121081 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAACA 121133


RESULT 11
AC003689
LOCUS       AC003689               137693 bp    DNA     linear   PRI 21-NOV-1998
DEFINITION  Homo sapiens Chromosome 11q12.2 PAC clone pDJ1081b4 containing
            human mRNA for T-cell glycoprotein CD6, complete sequence.
ACCESSION   AC003689
VERSION     AC003689.1  GI:3900834
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 137693)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 137693)
  AUTHORS   Evans,G.A., Athanasiou,M., Basit,M., Bradbury,P., Brignac,S.,
```

Bumesiter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwester Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE 3 (bases 1 to 137693)
AUTHORS Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Butler,C., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT On Nov 21, 1998 this sequence version replaced gi:2695567.
Further information regarding the map of this region or annotation of pDJ1081b4 can be found at http://gestec.swmed.edu/chromoso1.htm
IMPORTANT: This submission contains the entire insert of clone pDJ1081b4. pDJ1081b4 comes from the RPCI-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.98. In addition, attempts have been made to assure over 99% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.
CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome 11p12.2 Best's disease region mapped between STS D11S461 and EST AHNAK. This region spans over 1.5 Mbp.
MARKER CONFIRMATION: ESTs/STSs sequence confirmed; CD6, D11S461
MAPPED CLONE OVERLAP: PACs pDJ606g6 and pDJ59L21.

FEATURES Location/Qualifiers
source 1. .137693
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 164. .454
/rpt_family="Alu"
repeat_region complement(1543. .1835)
/rpt_family="Alu"
repeat_region complement(2058. .2360)
/rpt_family="Alu"
repeat_region complement(3515. .3796)
/rpt_family="Alu"
repeat_region complement(5954. .6244)
/rpt_family="Alu"
repeat_region complement(6962. .7215)
/rpt_family="Alu"
repeat_region complement(8088. .8263)
/rpt_family="Alu"
repeat_region complement(12551. .12719)
/rpt_family="Alu"
repeat_region 13480. .13676
/rpt_family="MER33"
repeat_region 13677. .13985
/rpt_family="Alu"
repeat_region complement(15815. .15950)
/rpt_family="Alu"
repeat_region complement(17641. .17934)
/rpt_family="Alu"
repeat_region 18423. .18693
/rpt_family="Alu"
repeat_region 18722. .19008
/rpt_family="Alu"
repeat_region complement(21315. .21594)
/rpt_family="Alu"
repeat_region complement(21636. .22481)
/rpt_family="MER42"
repeat_region complement(23396. .23697)
/rpt_family="Alu"
repeat_region 25083. .25349
/rpt_family="Alu"
repeat_region complement(28459. .28738)
/rpt_family="Alu"
repeat_region 29031. .29103
/rpt_family="Alu"
repeat_region complement(29491. .29785)
/rpt_family="Alu"
repeat_region complement(30052. .30172)
/rpt_family="Alu"
repeat_region complement(30261. .30640)
/rpt_family="Alu"
repeat_region 30885. .31193
/rpt_family="Alu"
repeat_region 32498. .32814
/rpt_family="Alu"
repeat_region complement(33478. .33782)
/rpt_family="Alu"
repeat_region 34395. .34556
/rpt_family="MIR"
repeat_region 35452. .35745
/rpt_family="Alu"
repeat_region complement(35739. .35983)
/rpt_family="MER33"
repeat_region complement(36141. .36244)
/rpt_family="MER3"
repeat_region complement(36206. .36244)
/rpt_family="MER33"
repeat_region 37244. .37491
/rpt_family="Alu"
repeat_region 38123. .38398
/rpt_family="Alu"
repeat_region complement(39547. .39837)
/rpt_family="Alu"
repeat_region 40124. .40270
/rpt_family="MER45"
repeat_region 40625. .40913
/rpt_family="Alu"
repeat_region complement(41015. .41075)
/rpt_family="MLT1"
repeat_region complement(41333. .41615)
/rpt_family="Alu"
repeat_region 41812. .42105
/rpt_family="Alu"
repeat_region 44026. .44309
/rpt_family="Alu"
repeat_region complement(47659. .47752)
/rpt_family="MER5"
repeat_region 50186. .50330
/rpt_family="MIR"
repeat_region 52843. .53128
/rpt_family="Alu"
repeat_region 53437. .53590
/rpt_family="MIR"
repeat_region 53688. .53967
/rpt_family="Alu"
repeat_region 56946. .57144
/rpt_family="Alu"
repeat_region 59439. .59731
/rpt_family="Alu"
repeat_region complement(61583. .61649)
/rpt_family="MIR"
repeat_region complement(62025. .62159)
/rpt_family="MIR"
repeat_region complement(62743. .63037)
/rpt_family="Alu"
repeat_region complement(63993. .64088)
/rpt_family="MIR"

```
     repeat_region   complement(66479. .66870)
                     /rpt_family="Alu"
     repeat_region   67416. .67584
                     /rpt_family="MER42"
     repeat_region   67594. .67887
                     /rpt_family="Alu"
     repeat_region   68119. .68182
                     /rpt_family="MER42"
     repeat_region   complement(69234. .69519)
                     /rpt_family="Alu"
     repeat_region   complement(69560. .69821)
                     /rpt_family="Alu"
     repeat_region   complement(70420. .70676)
                     /rpt_family="Alu"
     repeat_region   complement(70696. .70906)
                     /rpt_family="MER42"
     repeat_region   71247. .71537
                     /rpt_family="Alu"
     repeat_region   complement(71520. .71655)
                     /rpt_family="MER42"
     repeat_region   72778. .73053
                     /rpt_family="Alu"
     repeat_region   complement(75074. .75330)
                     /rpt_family="Alu"
     repeat_region   complement(76584. .76879)
                     /rpt_family="Alu"
     repeat_region   complement(76944. .77232)
                     /rpt_family="Alu"
     repeat_region   77719. .77994
                     /rpt_family="Alu"
     repeat_region   78902. .79179
                     /rpt_family="Alu"

  Query Match           22.9%;  Score 457.4;  DB 8;  Length 137693;
  Best Local Similarity   59.7%;  Pred. No. 2.8e-84;
  Matches  880;  Conservative    1;  Mismatches  577;  Indels   17;  Gaps    6;

Qy        398 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 456
              |||    | ||| ||| || || ||| ||||||| || |  ||| |||  |||||||  ||
Db      76572 TTTCTTTCTTTCTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 76631

Qy        457 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 516
              |  || |||  ||| |   | ||||  || ||||  ||||| | | ||||||| |   |
Db      76632 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 76691

Qy        517 GACCTCAGCCTCCCAAGTAGCTGGGACTACAGAGCGTGCACCACCACGCCTGGCTAATTAA 576
              |||||||||||| ||||||| || |||  || ||| |||||||||  ||| |||
Db      76692 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 76751

Qy        577 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 636
              |   || || |||||||| |||| |  | | ||||| |||||| ||||  | ||||||
Db      76752 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 76808

Qy        637 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 696
              | ||| |  |||| || |||| ||| ||||||||||||||||||| ||| ||||||||
Db      76809 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 76868

Qy        697 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 756
              || |  ||||  |  || |  |||    |      ||   ||    || |          |
Db      76869 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 76928

Qy        757 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 816
              |      ||     || |||| ||       | ||           |   ||  |   |
Db      76929 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 76988

Qy        817 TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 876
              |      || |||||||||| |||||||||  ||||||  |:||||||||||||||
Db      76989 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCT 77048

Qy        877 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--G 934
              ||||||| |||||||||| |||||| | | || |||||| |  |||||| ||||||
Db      77049 CAGCCTCCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGT 77108

Qy        935 TATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCT 994
              ||||||| |||||  |||||| |||||||||  |||||||||||| ||||||||||||
Db      77109 ATTTTTTAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 77168

Qy        995 CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTG 1054
              || |||||| |||||||||| || |||||||||||| |||| |||| |||| ||| | |
Db      77169 CAGGTGATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCG 77228

Qy       1055 CCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1114
              || |    |  |  ||    |  |||||    |  |    |  || |
Db      77229 CCCGATGCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TA 77287

Qy       1115 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1174
              ||| |  |  |  ||          |   | |          || |   | || |
Db      77288 TCATTTCCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATG 77347

Qy       1175 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1234
              || || ||   | | ||     |  | | ||   | || | |   |  ||| | | |
Db      77348 ATTATTATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATT 77407

Qy       1235 TTTTCA---------CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTAT 1285
              | |||          | || | |  |  |  |||   | |||||        | |
Db      77408 ATATCAATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAA 77467

Qy       1286 TGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTA 1345
              ||    | |  ||  |  |  | ||  | ||  |  | |        |   |
Db      77468 TGGGTGTTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGT 77527

Qy       1346 TACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACC 1405
              | |   | |        |||| |  |   || |  | |   |  | | |   |
Db      77528 TTCTAGAACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGT 77587

Qy       1406 CTCTCCCCAGAGAAATACACACACACACACACTCACACACAGT-TTTTTTTTAATGTT 1464
              | |    ||||  |      | |  |  |  |  ||  | |     ||   |
Db      77588 CCATGGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGC 77647

Qy       1465 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1524
              || |  ||||  |          |  | ||| ||  | || |  |  |
Db      77648 CAAAAATGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGA 77707

Qy       1525 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1584
              |   ||   ||| |||| |  ||||| |  || |||||||||| || |||  |||||
Db      77708 AGGCCAGGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGG 77767

Qy       1585 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1644
              || ||||||||| || |||||||||||||||||| ||| |||||||||| || |||||||
Db      77768 ATTACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTAC 77827

Qy       1645 TAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 1704
              |||||||||||||||||||||||||| || ||||||| ||||||| |||||||||| ||||
Db      77828 TAAAAATACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGA 77887

Qy       1705 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 1764
              |||||||||  ||||| ||||||||| ||||||| |||||||||||||||||||||||
Db      77888 GGCTGAGGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGT 77947

Qy       1765 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATA 1824
              |||||||||||||||||||| ||| | | ||| ||| |||||||| ||  ||||||||
Db      77948 GTCACTGCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATA 78007

Qy       1825 AAATAAAGGATCGGAGAGAAACAAAACTAATAAGA 1859
              |||  ||  ||    |   |||  |   |||  | |
Db      78008 AAAATAATGAAAATAATAAAATGATGTTAAAGATA 78042
```

```
RESULT 12
AP003721
LOCUS       AP003721              201460 bp    DNA     linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-881M11, complete
            sequence.
ACCESSION   AP003721
VERSION     AP003721.3  GI:31790667
KEYWORDS    HTG.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 201460)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 16, 2003 this sequence version replaced gi:28189463.
FEATURES             Location/Qualifiers
     source          1. .201460
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-881M11"
ORIGIN

  Query Match             22.8%;  Score 456.2;  DB 8;  Length 201460;
  Best Local Similarity   59.7%;  Pred. No. 4.8e-84;
  Matches  877;  Conservative    1;  Mismatches  574;  Indels   17;  Gaps    6;

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAG 462
Db     144460 TTTCTTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGAGTGCAA 144519

Qy        463 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 522
Db     144520 TGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCCTGCCTC 144579

Qy        523 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 582
Db     144580 AGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTTGTA--- 144636

Qy        583 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 642
Db     144637 TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTGACCTCA 144696

Qy        643 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 702
Db     144697 GGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCATCACGCC 144756

Qy        703 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 762
Db     144757 CGGCCTCACATATTCTTATCCTAATATGTTTTCACGCTTTGAAGTCCTACAGTAATCATC 144816

Qy        763 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 822
Db     144817 CTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCAATGGCG 144876

Qy        823 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 882
Db     144877 CAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCTCAGCCT 144936

Qy        883 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--GTATTTT 940
Db     144937 CCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGTATTTTT 144996

Qy        941 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1000
Db     144997 TAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTG 145056

Qy       1001 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1060
Db     145057 ATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGAT 145116

Qy       1061 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1120
Db     145117 GCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TATCATTT 145175

Qy       1121 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1180
Db     145176 CCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATGATTATT 145235

Qy       1181 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1240
Db     145236 ATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATTATATCA 145295

Qy       1241 C---------TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTAC 1291
Db     145296 ATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAATGGGTG 145355

Qy       1292 TGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATC 1351
Db     145356 TTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGTTTCTAG 145415

Qy       1352 CTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCC 1411
Db     145416 AACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGTCCATGG 145475

Qy       1412 CCAGAGAAATACACACACACACACACTCACACACAGT-TTTTTTTTAATGTTTGCAAC 1470
Db     145476 AGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGCCAAAAA 145535

Qy       1471 TAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTG 1530
Db     145536 TGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGAAGGCCA 145595

Qy       1531 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 1590
Db     145596 GGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGGATTACT 145655

Qy       1591 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 1650
Db     145656 TGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTACTAAAAA 145715

Qy       1651 TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGA 1710
Db     145716 TACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGAGGCTGA 145775

Qy       1711 GGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACT 1770
Db     145776 GGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGTCACT 145835

Qy       1771 GCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 1830
Db     145836 GCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATAAAAATA 145895

Qy       1831 AGGATCGGAGAGAAACAAAACTAATAAGA 1859
Db     145896 ATGAAAATAATAAAATGATGTTAAAGATA 145924


RESULT 13
AC116933
LOCUS       AC116933              195986 bp    DNA     linear   HTG 21-DEC-2002
DEFINITION  Papio anubis clone RP41-1C24, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC116933
VERSION     AC116933.2  GI:27356688
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 195986)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Dec 21, 2002 this sequence version replaced gi:19909407.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cqw
                 Center clone name: 001C24

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 195122 bases at least Q40
                 Consensus quality: 195370 bases at least Q30
                 Consensus quality: 195594 bases at least Q20
                 Insert size: 186000; agarose-fp
                 Insert size: 195686; sum-of-contigs
                 Quality coverage: 9.19x in Q20 bases; agarose-fp
                 Quality coverage: 8.74x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    76152: contig of 76152 bp in length
            *    76153    76252: gap of unknown length
            *    76253   126668: contig of 50416 bp in length
            *   126669   126768: gap of unknown length
            *   126769   177394: contig of 50626 bp in length
            *   177395   177494: gap of unknown length
            *   177495   195986: contig of 18492 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195986
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-1C24"
                     /clone_lib="RP41"
     misc_feature    1. .76152
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .54626
                     /note="clone overlaps with GenBank Accession Number
                     AC116934 clone RP41-206F5 (center project name cqv)"
     gap             76153. .76252
                     /estimated_length=unknown
     misc_feature    76253. .126668
                     /note="assembly_fragment"
     gap             126669. .126768
                     /estimated_length=unknown
     misc_feature    126769. .177394
                     /note="assembly_fragment"
     gap             177395. .177494
                     /estimated_length=unknown
     misc_feature    177495. .195986
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             22.6%;  Score 451.2;  DB 14;  Length 195986;
  Best Local Similarity   61.1%;  Pred. No. 5.1e-83;
  Matches  891;  Conservative    1;  Mismatches  514;  Indels   52;  Gaps    8;

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
              |||| ||| || ||   ||||||||| ||||||||||||||||||||  |||  ||||
Db     180502 TTTTCCTTTTTTTTTAAGTAGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGT 180561

Qy        464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
               ||||||   |   || ||||| |||  |  ||||| |  || | ||||  |  ||||||
Db     180562 AGTGCAATAGTGTTTCCCTGCATCCTTCAATTCCTGGGAACAGGTGATCCTCCCACCTCA 180621

Qy        524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
              | || |  |||||||||||||    ||      |||||| |||  ||||| | ||||||||
Db     180622 GTCTTCTGAGTAGCTGGGACTTTAGGCACATGCCACCATGCCCAGCTAACTTAAAAAATT 180681

Qy        584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643
              |||||| ||| || |||||||  ||  ||||  ||||||  |||  ||||  | |||  ||||
Db     180682 TTTTTGAAGACACAGGGTCTTGCTCTGTTGCTCAGGCTACTCTCAAACAGCCGGCCTCAA 180741

Qy        644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703
              |  ||||||   |||   | |||||||||||||||||| ||| ||||||||||||  |
Db     180742 GTGATCCTCTCGCCTCATCCTCCCAAAGTGCTGGGATTAGAGGTGTGAGCCACCATGCCC 180801

Qy        704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 763
               ||   || ||| || |  |   |  ||          | |||  |  |          |
Db     180802 AGCCCATTTTTTTTTAATTTCTTAATTTAATTTCTATTTTTAATTTTTA---TTATTTTA 180858

Qy        764 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 823
              |||  |     |||  ||  ||   |||||||||||||||||||| | ||  |     |||
Db     180859 TTATTTTTATTTTATTTATTTATATTTTTGAGACAGAGTCTCACTCCGTCGCACAGGCTG 180918

Qy        824 CCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCT 869
                ||              |||||||||||||||||| ||||||||||||||||: | ||
Db     180919 GAGTGCAATGGTGTGATCTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCAACTCC 180978

Qy        870 CCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATT 929
              ||||||| |||||||||||||||||| |||||| | |  | |||| |  | ||||||| |
Db     180979 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGGCCTCCACTACACCTGGCTAA-T 181037

Qy        930 TTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCT 989
```

```
Db     181038 TTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGTCAGGCTGGTCTTGAACTCCT 181097

Qy        990 GACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCA 1049

Db     181098 GACCTCAGGTGACCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA 181157

Qy       1050 CTGTGCCTGGCCGATTACTGTCTATTTTCT----------TTATTGCTATATCCCCAGAT 1099

Db     181158 CTGTGCCCAGCAACCCCATTTCTTTTTTTTAAACAAATGCTTACTAACCTACTCTAAGAG 181217

Qy       1100 CTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTA 1159

Db     181218 CTGTAAGGGCAAATGAGGCAATGTCTTTACCCTCAGAATCTTATCTTTTGGTGAGGAGAT 181277

Qy       1160 GATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACA 1219

Db     181278 AATACACATATATAAATAATGATACAAGGACATATATGATAAATGCCATGACAGATTATA 181337

Qy       1220 ATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAAC 1279

Db     181338 AATCCATATTCAC-----TACACAAACAAGAGGCTTAGAAAAGTTCAGATTCCCACTTTG 181392

Qy       1280 CTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATT 1339

Db     181393 GGAGGCCGAGACAGGTGGATCACCTGAGGTCAGGAGTTCA--AGAGCAGCGTGGCCAACA 181450

Qy       1340 TGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTA 1399

Db     181451 TGGTGAAACTCCGTCTCTACTAAAAGTACAAAAATTAGCCGGGGGTAGTGGAGGGCACCT 181510

Qy       1400 TGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTA 1459

Db     181511 GTAATCCCAGCTACTCAGAAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGG 181570

Qy       1460 ATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGT---------------T 1504

Db     181571 CTGCAGTGAGCCGAGATTGCACCACTGCACTCTAGCCTGGGGGACAGAGCCAGACTCCAT 181630

Qy       1505 CATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTT 1564

Db     181631 CTCAAAAAAAAAAAAAAAAAAGGCCGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTT 181690

Qy       1565 TGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATA 1624

Db     181691 TAGGAGCCTGAGGCGGGCAGATCA--TGAGGTCAGGAGTTCGAGACCAGCCTGGCCAATA 181748

Qy       1625 TGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCT 1684

Db     181749 TGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGCCACGCACCT 181808

Qy       1685 GTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGG 1744

Db     181809 GTAGTCCCAGCTACTCAAGAGGCCGAGGCAGGAGAATCACCTGAACCCAGGAGGTGGAGG 181868

Qy       1745 TTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTAT 1804

Db     181869 TTGCAGTGAGCCGAAATGGTGCCATTGCACTCCAGCCTGGGTGTCAGAGCGAGACTCCAT 181928

Qy       1805 CTCAAAAAAATAAATAAA 1822

Db     181929 CTCAAAAAGAAAAAGAAA 181946


RESULT 14
AC005052/c
LOCUS       AC005052              134210 bp    DNA     linear   PRI 08-NOV-2000
DEFINITION  Homo sapiens BAC clone CTB-38K21 from Xq23, complete sequence.
ACCESSION   AC005052
VERSION     AC005052.2  GI:10122134
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 134210)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 134210)
  AUTHORS   Tin-Wollam,A., Graves,T. and Cofman,M.
  TITLE     The sequence of Homo sapiens BAC clone CTB-38K21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 134210)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 14, 2000 this sequence version replaced gi:3212939.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_RG038K21
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the chromosome X mapping group
            at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/

            SOURCE INFORMATION:
            Clone CTB-38K21 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP3-327A19, 200 base pair
            overlap.  Actual start of this clone is at base position 1 of
            CTB-38K21; actual end is at base position 9416 of RP3-327A19.
```

```
FEATURES             Location/Qualifiers
     source          1. .134210
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq23"
                     /clone="CTB-38K21"
                     /clone_lib="CITB-978SK-B"
     repeat_region   20. .268
                     /rpt_family="Alu"
     repeat_region   269. .561
                     /rpt_family="Alu"
     repeat_region   978. .1285
                     /rpt_family="Alu"
     misc_feature    2964. .4303
                     /note="CpG_island (%GC=68.3, o/e=0.70, #CpGs=110)"
     misc_feature    3597. .3895
                     /note="similar to  EST BE299003 (NID:g9182751)"
     repeat_region   6613. .6919
                     /rpt_family="Alu"
     repeat_region   7154. .7214
                     /rpt_family="L2"
     repeat_region   7640. .7924
                     /rpt_family="Alu"
     repeat_region   8152. .8333
                     /rpt_family="MER1_type"
     repeat_region   8565. .8749
                     /rpt_family="MaLR"
     repeat_region   8750. .9044
                     /rpt_family="Alu"
     repeat_region   9045. .9256
                     /rpt_family="MaLR"
     repeat_region   9271. .9700
                     /rpt_family="L2"
     repeat_region   10238. .10290
                     /rpt_family="L2"
     repeat_region   10291. .10585
                     /rpt_family="Alu"
     repeat_region   10586. .10666
                     /rpt_family="L2"
     repeat_region   10667. .10822
                     /rpt_family="Alu"
     repeat_region   10823. .11015
                     /rpt_family="L2"
     repeat_region   11016. .11199
                     /rpt_family="MER1_type"
     repeat_region   11200. .11309
                     /rpt_family="L2"
     repeat_region   11310. .11544
                     /rpt_family="Alu"
     repeat_region   11545. .11610
                     /rpt_family="L2"
     repeat_region   11860. .12166
                     /rpt_family="Alu"
     misc_feature    12126. .12175
                     /note="similar to  EST AI025436 (NID:g3241049) ow27h07.s1"
     repeat_region   12178. .12228
                     /rpt_family="L1"
     repeat_region   12405. .12520
                     /rpt_family="L1"
     repeat_region   12712. .13008
                     /rpt_family="Alu"
     repeat_region   13009. .13315
                     /rpt_family="Alu"
     repeat_region   13603. .13850
                     /rpt_family="ERVL"
     repeat_region   13851. .14147
                     /rpt_family="Alu"
     repeat_region   14148. .14222
                     /rpt_family="ERVL"
     repeat_region   14223. .14526
                     /rpt_family="Alu"
     repeat_region   14527. .14713
                     /rpt_family="ERVL"
     repeat_region   14815. .15244
                     /rpt_family="L2"
     repeat_region   15245. .15566
                     /rpt_family="Alu"
     repeat_region   15567. .16446
                     /rpt_family="L2"
     repeat_region   16447. .16580
                     /rpt_family="Alu"
     repeat_region   16581. .16866
                     /rpt_family="L2"
     repeat_region   16867. .17178
                     /rpt_family="Alu"
     repeat_region   17179. .17251
                     /rpt_family="L2"
     repeat_region   17255. .17390
                     /rpt_family="MER1_type"
     repeat_region   17872. .18165
                     /rpt_family="Alu"
     repeat_region   18239. .18532
                     /rpt_family="Alu"
     repeat_region   18540. .18841
                     /rpt_family="Alu"
     misc_feature    18631. .18876
                     /note="similar to  EST AW867673 (NID:g8001829)"
     repeat_region   19286. .19570
                     /rpt_family="L2"
     repeat_region   19885. .20185
                     /rpt_family="Alu"
     repeat_region   21149. .21446
                     /rpt_family="Alu"
     repeat_region   21559. .21803
                     /rpt_family="Alu"
     repeat_region   21835. .21951
                     /rpt_family="L1"
     repeat_region   21979. .22286
                     /rpt_family="Alu"
     repeat_region   23260. .23565
                     /rpt_family="Alu"
     repeat_region   23568. .23864
                     /rpt_family="Alu"
     misc_feature    23875. .23924
                     /note="similar to  EST AI735654 (NID:g5057178) at19b08.x1"
     misc_feature    23875. .23924
                     /note="similar to  EST AW619972 (NID:g7326156)"
     misc_feature    23878. .23926

  Query Match             22.5%;  Score 449.2;  DB 8;  Length 134210;
  Best Local Similarity   59.5%;  Pred. No. 1.4e-82;
  Matches  896;  Conservative    1;  Mismatches  554;  Indels   55;  Gaps    6;

Qy        355 TGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGT 414
              || | |||||||   |  |          | |   |||| |  |  ||    ||| ||| |
Db      53793 TGTGGTATTTGCTTTATTTATTTATTTATTTTTTTGGTTAGGGGGTGCTGATTTCTTTTT 53734

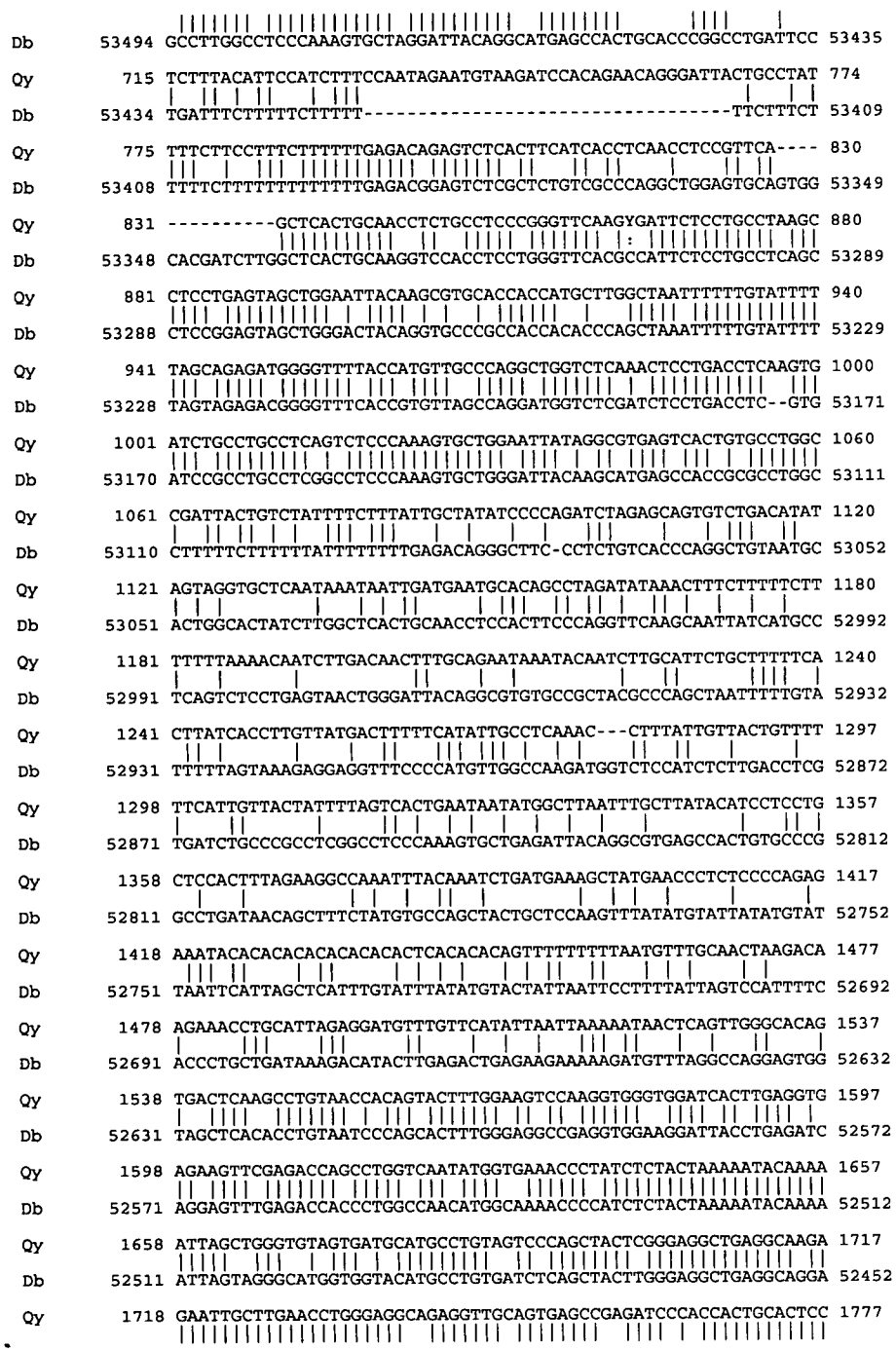
Qy        415 TGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCAT 474
              | || ||| |||||||||||||| ||||||||||||||||  |||  |||||   | | |||
Db      53733 TCTTTTTTTGAGACAGGGTCTTTCTCTGTCACCCAGGCTGGAGTGCAGTGACAC-ATCAT 53675

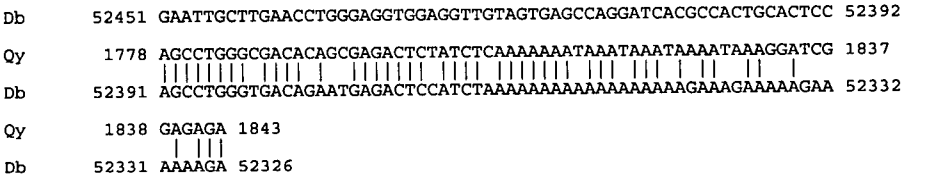
Qy        475 AGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGT 534
              | |||||||||||||||||||||| |||||||  |||  |  | ||||||||||| |||
Db      53674 GGCTCACTGCAGCCTCAACCTCCTGGGCTCAAGCAATCCTCAAATCTCAGCCTCCCGAGT 53615

Qy        535 AGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAG 594
              ||||||| |||   || ||  | |||| ||| ||||||||   ||| | ||||
Db      53614 AGCTGGGTCTATAGGCTTGAGCTACCAGGCCCGGCTAATTTTTAAATTGTTTTGTAGAGA 53555

Qy        595 ACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCT 654
              ||||||  |||| ||   ||||| |||||||||||||||| ||||||  ||||||||
Db      53554 CAAGGGTCTCACTATGTCACCCAGGGTTGTCTTAAACTCCTGGGTTCAAGGGATCCTCCT 53495

Qy        655 ACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATT 714
```





RESULT 15
AC005484/c
LOCUS AC005484 131943 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP5-847O8 from 14q24.3, complete sequence.
ACCESSION AC005484
VERSION AC005484.2 GI:5091654
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131943)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 9847074
REFERENCE 2 (bases 1 to 131943)
AUTHORS Cloud,J., Wohldmann,P. and Holmes,A.
TITLE The sequence of Homo sapiens PAC clone RP5-847O8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131943)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 131943)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 131943)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jun 17, 1999 this sequence version replaced gi:3907511.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ0847O8
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

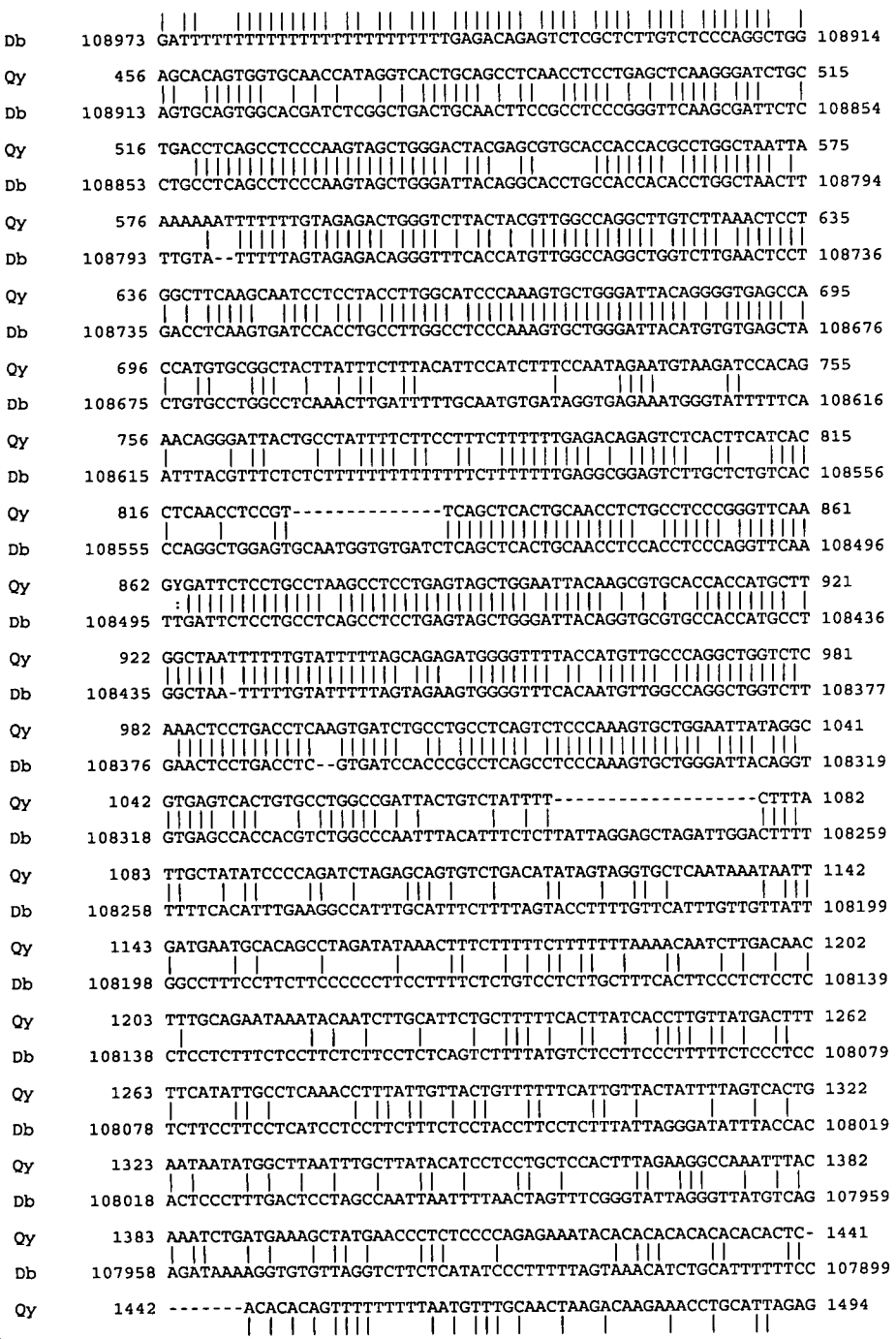
This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

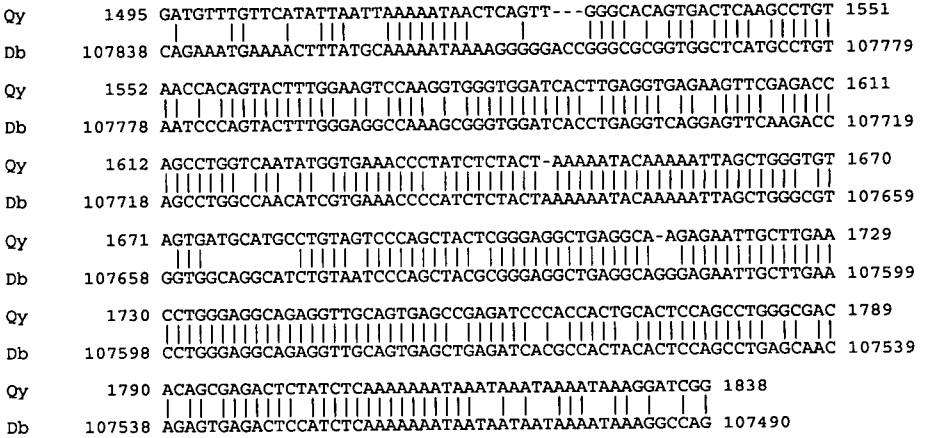
SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-892G5; the clone sequenced to the right is RP4-592G7, 200 bp overlap. Actual start of this clone is at base position 1 of RP5-847O8; actual end is at base position 131747 of RP5-847O8.

FEATURES Location/Qualifiers
source 1. .131943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/clone="RP5-847O8"
/clone_lib="RPCI-5"
repeat_region 1058. .1098
/rpt_family="AT_rich"
repeat_region 1762. .2225
/rpt_family="MaLR"
repeat_region 2305. .2587
/rpt_family="Alu"
repeat_region 2937. .3086
/rpt_family="L2"
repeat_region 3293. .3591
/rpt_family="Alu"
repeat_region 3651. .3916
/rpt_family="L1"
repeat_region 3917. .3943
/rpt_family="(TA)n"
repeat_region 4815. .4841
/rpt_family="AT_rich"
repeat_region 5153. .5244
/rpt_family="L1"
repeat_region 5728. .6008
/rpt_family="Alu"
repeat_region 6009. .6034
/rpt_family="(CAA)n"
misc_feature 8080. .8468
/note="match to EST R05271 (NID:g755891) ye91b03.s1"
STS 8080. .8277
repeat_region 8284. .8589
/rpt_family="Alu"
misc_feature 8649. .9039
/note="match to EST R05328 (NID:g755948) ye91b03.r1"
repeat_region 8857. .9007
/rpt_family="MIR"
repeat_region 9062. .9152
/rpt_family="MIR"
repeat_region 9216. .9288
/rpt_family="MIR"
repeat_region 9476. .9770
/rpt_family="Alu"
STS 9711. .10034
repeat_region 10004. .10114
/rpt_family="Alu"
repeat_region 10115. .10135
/rpt_family="(TAAA)n"
repeat_region 10136. .10158
/rpt_family="Alu"
repeat_region 10205. .10501
/rpt_family="Alu"
repeat_region 10513. .10819
/rpt_family="Alu"
repeat_region 10890. .11242
/rpt_family="MaLR"
repeat_region 11246. .11519
/rpt_family="Alu"
repeat_region 11567. .11996
/rpt_family="L2"
repeat_region 12006. .12113
/rpt_family="L2"
repeat_region 12123. .12146
/rpt_family="(CAAAA)n"
repeat_region 12147. .12430
/rpt_family="Alu"
repeat_region 12445. .12535
/rpt_family="L1"
repeat_region 12527. .13097
/rpt_family="L1"
repeat_region 13217. .13339
/rpt_family="Alu"
repeat_region 13389. .13994
/rpt_family="L2"
repeat_region 14138. .14158
/rpt_family="AT_rich"
repeat_region 14195. .15071
/rpt_family="L1"
repeat_region 15048. .15521
/rpt_family="L1"
repeat_region 15520. .15992
/rpt_family="L1"
repeat_region 15989. .16221
/rpt_family="L1"
repeat_region 16257. .16556
/rpt_family="Alu"
repeat_region 16569. .16963
/rpt_family="L1"
repeat_region 17016. .17039
/rpt_family="AT_rich"
repeat_region 17073. .17182
/rpt_family="MIR"
repeat_region 17252. .17452
/rpt_family="MaLR"
repeat_region 17473. .17991
/rpt_family="L2"
repeat_region 18106. .18192
/rpt_family="MER1_type"
repeat_region 18193. .18504
/rpt_family="Alu"
repeat_region 18505. .18588
/rpt_family="MER1_type"
repeat_region 18624. .18904
/rpt_family="L1"
repeat_region 19144. .19368
/rpt_family="Alu"
repeat_region 20030. .20159
/rpt_family="Alu"
repeat_region 20167. .20317
/rpt_family="Alu"
repeat_region 20318. .20617
/rpt_family="Alu"
repeat_region 20618. .20753
/rpt_family="Alu"
repeat_region 20802. .20954
/rpt_family="L1"
repeat_region 20955. .21256
/rpt_family="Alu"

Query Match 22.4%; Score 447.8; DB 8; Length 131943;
Best Local Similarity 61.3%; Pred. No. 2.7e-82;
Matches 913; Conservative 1; Mismatches 523; Indels 52; Gaps 10;

Qy 397 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATG 455





Search completed: February 9, 2006, 18:20:35
Job time : 10781.2 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:16:14 ; Search time 1377.77 Seconds
(without alignments)
9679.442 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 1998.6
Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1998.6	100.0	12174	12	ADI35082	Adi35082 Human PLA
2	1998.6	100.0	12174	12	ADJ09983	Adj09983 Human pho
3	1998.6	100.0	13612	6	ABK47376	Abk47376 Human Pho
4	434.2	21.7	93544	13	ABD33504	Abd33504 Human can
c 5	434.2	21.7	160482	11	ACN43914	Acn43914 Human gen
c 6	433.2	21.7	91760	11	ACN44410	Acn44410 Human gen
c 7	425.4	21.3	38538	13	ABD33357	Abd33357 Human can
8	417.2	20.9	122888	6	ABK83569	Abk83569 Human cDN
c 9	416.2	20.8	6405	5	AAF97850	Aaf97850 Human neu
c 10	416	20.8	126990	12	ADP13332	Adp13332 Renal cel
c 11	414.2	20.7	196686	11	ACN44170	Acn44170 Human gen
c 12	413.6	20.7	86000	12	ADP68568	Adp68568 Human PPA
c 13	413.6	20.7	215974	12	ADQ97523	Adq97523 Human can
c 14	413.4	20.7	13670	6	AAL42369	Aal42369 Human Gua
15	413.2	20.7	6519	5	ABA16909	Aba16909 Human ner
16	408.8	20.5	17245	4	AAK83897	Aak83897 Human imm
17	406.2	20.3	348101	12	ADQ97146	Adq97146 Human can
c 18	405.8	20.3	227246	13	ABD33272	Abd33272 Human can
c 19	404	20.2	95240	10	ADL13556	Adl13556 Osteoarth
c 20	403.6	20.2	23639	6	ABQ78991	Abq78991 Human pho
c 21	400.4	20.0	14176	4	AAS26670	Aas26670 Human gen
c 22	400.4	20.0	14176	8	ABX74019	Abx74019 Human nov
c 23	400.4	20.0	177531	8	ACF62732	Acf62732 Cancer ba
c 24	400.4	20.0	177531	8	ADB20847	Adb20847 MRP1 base
c 25	400.4	20.0	177531	10	ADB87936	Adb87936 Human UGT
c 26	400.4	20.0	177531	10	ADB96919	Adb96919 Human MDR
c 27	400.4	20.0	177531	10	ADB92110	Adb92110 Human MDR
c 28	400.4	20.0	177531	10	ADH74617	Adh74617 Human BAC
29	399	20.0	87687	11	ACN45166	Acn45166 Human gen
30	399	20.0	131078	14	ADX06911	Adx06911 Cyclin-de
31	398.6	19.9	174448	11	ACN43946	Acn43946 Human gen
c 32	398.4	19.9	167343	6	ABL64403	Abl64403 Stomach c
c 33	398.4	19.9	167343	6	ABL67239	Abl67239 Thyroid c
34	394.8	19.8	30620	4	AAK66931	Aak66931 Human imm
35	394.6	19.7	19300	12	ADP74371	Adp74371 Human X c
36	392.4	19.6	9289	10	ADE84033	Ade84033 5' regula
37	391.4	19.6	226475	9	AAD58279	Aad58279 Human tum
c 38	390	19.5	13224	4	AAS41751	Aas41751 Genomic s
c 39	390	19.5	13224	4	ABA06811	Aba06811 Human gen
c 40	390	19.5	13224	6	ABV84148	Abv84148 Human pol
41	390	19.5	36221	4	AAS00624	Aas00624 Human dea
c 42	389.6	19.5	23456	13	ABD33110	Abd33110 Human can
43	389.6	19.5	60057	11	ACN44314	Acn44314 Human gen
44	388	19.4	75252	11	ACN44450	Acn44450 Human gen
45	387.8	19.4	48001	12	ADO61194	Ado61194 Human PPA

ALIGNMENTS

RESULT 1
ADI35082
ID ADI35082 standard; DNA; 12174 BP.
XX
AC ADI35082;
XX
DT 22-APR-2004 (first entry)
XX
DE Human PLA2G1B nucleotide sequence.
XX
KW PLA2G1B ; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC subject, where the presence of the polymorphic variation indicates a
CC predisposition to fat deposition in the subject. The polymorphic
CC variation is a guanine at position 7328 or thymine at position 9182 of
CC the present sequence. The method is useful for diagnosing a
CC predisposition to fat deposition or leanness in a subject, and
CC consequently for diagnosing a predisposition to non-insulin dependent
CC diabetes mellitus (NIDDM) in a subject and conditions such as
CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC hypercoagulability, or microalbuminuria, which can lead to early
CC prescription of preventive measures. The present sequence represents a
CC human PLA2G1B nucleotide sequence.
XX
SQ Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

Query Match 100.0%; Score 1998.6; DB 12; Length 12174;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
Db    7200 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 7259

Qy      61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
Db    7260 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 7319

Qy     121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
Db    7320 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 7379

Qy     181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
Db    7380 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 7439

Qy     241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
Db    7440 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 7499

Qy     301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
Db    7500 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 7559

Qy     361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 420
Db    7560 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 7619

Qy     421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
Db    7620 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 7679

Qy     481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
Db    7680 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 7739

Qy     541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
Db    7740 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 7799

Qy     601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
Db    7800 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 7859

Qy     661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
Db    7860 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 7919

Qy     721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
Db    7920 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 7979

Qy     781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
Db    7980 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 8039

Qy     841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
Db    8040 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 8099

Qy     901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
Db    8100 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 8159

Qy     961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
Db    8160 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 8219

Qy    1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
Db    8220 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 8279

Qy    1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
Db    8280 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 8339

Qy    1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200
Db    8340 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 8399

Qy    1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
Db    8400 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 8459

Qy    1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
Db    8460 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 8519

Qy    1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
Db    8520 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 8579

Qy    1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
Db    8580 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 8639

Qy    1441 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
Db    8640 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 8699

Qy    1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
Db    8700 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8759

Qy    1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
Db    8760 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8819

Qy    1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
Db    8820 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8879

Qy    1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
Db    8880 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8939

Qy    1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
Db    8940 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8999

Qy    1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
Db    9000 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 9059

Qy    1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
Db    9060 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 9119
```

```
Qy       1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9120 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 9179

Qy       1981 ACKCTACATAAATATCTATTA 2001
              |||||||||||||||||||||
Db       9180 ACKCTACATAAATATCTATTA 9200
```

RESULT 2
ADJ09983
ID ADJ09983 standard; DNA; 12174 BP.
XX
AC ADJ09983;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(436,c)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(839,a)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(4050,a)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(4689,t)
FT		/*tag= d
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6282,a)
FT		/*tag= e
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6358,c)
FT		/*tag= f
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(6653,t)
FT		/*tag= g
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(7256,t)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(7300,a)
FT		/*tag= i
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(7301,c)
FT		/*tag= j
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(7328,g)
FT		/*tag= k
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(8062,c)
FT		/*tag= l
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(9182,t)
FT		/*tag= m
FT		/standard_name= "Single nucleotide polymorphisms"
FT	variation	replace(11649,c)
FT		/*tag= n
FT		/standard_name= "Single nucleotide polymorphisms"

XX
PN WO2004002296-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020831.
XX
PR 27-JUN-2002; 2002US-0392362P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
PI Rubin B;
XX
DR WPI; 2004-071944/07.
DR P-PSDB; ADJ09984.
XX
PT Identifying a candidate therapeutic for fat reduction, useful for
PT treating diabetes, by introducing a test molecule to a system comprising
PT PLA2G1B protein or nucleic acid, and determining the presence of
PT interaction between the compounds.
XX
PS Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC This invention relates to a novel candidate therapeutic agent useful for
CC fat reduction and disorders related to fat depositions. Specifically, it
CC refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC which is located on chromosome 12q24 and has been associated with central
CC fat deposition. The present invention describes methods to detect the
CC presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC in particular G7328A and T9182G, and subsequently provide treatment that
CC reduces fat deposition. This treatment may consist of an appetite
CC suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC deposition and it can be used to treat both obesity and non-insulin
CC dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC such as hypertension. As such, it exhibits antidiabetic activity. This
CC polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

Query Match 100.0%; Score 1998.6; DB 12; Length 12174;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1995; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       7200 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 7259

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              ||||||||||||||||||||||||||||||||||||||||||::||||||||||||||||||
Db       7260 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 7319

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7320 TCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 7379

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7380 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 7439

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7440 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 7499

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7500 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 7559

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7560 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 7619

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
```

```
Db      7620 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 7679
Qy       481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
Db      7680 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 7739
Qy       541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
Db      7740 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 7799
Qy       601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
Db      7800 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 7859
Qy       661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
Db      7860 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 7919
Qy       721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
Db      7920 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 7979
Qy       781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
Db      7980 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 8039
Qy       841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
Db      8040 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 8099
Qy       901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
Db      8100 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 8159
Qy       961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
Db      8160 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 8219
Qy      1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
Db      8220 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 8279
Qy      1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
Db      8280 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 8339
Qy      1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 1200
Db      8340 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACA 8399
Qy      1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
Db      8400 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 8459
Qy      1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
Db      8460 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 8519
Qy      1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
Db      8520 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 8579
Qy      1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
Db      8580 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 8639
Qy      1441 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
Db      8640 CACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 8699
Qy      1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
Db      8700 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8759
Qy      1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
Db      8760 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8819
Qy      1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
Db      8820 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8879
Qy      1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
Db      8880 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8939
Qy      1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
Db      8940 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8999
Qy      1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
Db      9000 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 9059
Qy      1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
Db      9060 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 9119
Qy      1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
Db      9120 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 9179
Qy      1981 ACKCTACATAAATATCTATTA 2001
Db      9180 ACTCTACATAAATATCTATTA 9200
```

RESULT 3
ABK47376
ID ABK47376 standard; DNA; 13612 BP.
XX
AC ABK47376;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(3845,A)
FT		/*tag= a
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	variation	replace(3968,A)
FT		/*tag= b
FT		/label= SNP
FT		/note= "Single nucleotide polymorphism"
FT	CDS	4053. .9613
FT		/*tag= c
FT		/product= "Pla2G1B"
FT	exon	4053. .4086
FT		/*tag= d
FT		/number= 1
FT	intron	4087. .5785
FT		/*tag= e
FT		/number= 1
FT	exon	5786. .5945
FT		/*tag= f
FT		/number= 2
FT	intron	5946. .6744

```
FT                   /*tag=  g
FT                   /number= 2
FT   variation       replace(6060,A)
FT                   /*tag=  h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   exon            6745. .6873
FT                   /*tag=  i
FT                   /number= 3
FT   variation       replace(6844,A)
FT                   /*tag=  j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   intron          6874. .9489
FT                   /*tag=  k
FT                   /number= 3
FT   exon            9490. .9613
FT                   /*tag=  l
FT                   /number= 4
FT   variation       replace(9531,A)
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1998.6;  DB 6;  Length 13612;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1995;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db       6716 CTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTATGAC 6775

Qy         61 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTAT 120
              |||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db       6776 CAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAACCCGTACACCCACACCTAT 6835

Qy        121 TCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 180
              ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6836 TCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACC 6895

Qy        181 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6896 TATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 6955

Qy        241 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6956 TTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 7015

Qy        301 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7016 GGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCT 7075

Qy        361 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7076 ATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGT 7135

Qy        421 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7136 TTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCA 7195

Qy        481 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7196 CTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGG 7255

Qy        541 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7256 GACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGG 7315

Qy        601 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7316 TCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTG 7375

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7376 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 7435

Qy        721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7436 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 7495

Qy        781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7496 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCA 7555

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
              ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db       7556 ACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 7615

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 7675

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7676 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 7735

Qy     1021 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7736 CAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTT 7795

Qy     1081 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7796 TATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAA 7855

Qy     1141 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTAAAACAATCTTGACA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7856 TTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTAAAACAATCTTGACA 7915

Qy     1201 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7916 ACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACT 7975

Qy     1261 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7976 TTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCAC 8035

Qy     1321 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8036 TGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTT 8095

Qy     1381 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8096 ACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACT 8155

Qy     1441 CACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8156 CACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTT 8215

Qy     1501 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8216 TGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGT 8275

Qy     1561 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8276 ACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTC 8335

Qy     1621 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8336 AATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCAT 8395

Qy     1681 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8396 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCA 8455

Qy     1741 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8456 GAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACT 8515

Qy     1801 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8516 CTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGAT 8575

Qy     1861 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8576 TCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACT 8635

Qy     1921 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8636 GTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAA 8695

Qy     1981 ACKCTACATAAATATCTATTA 2001
            ||:||||||||||||||||||
Db     8696 ACTCTACATAAATATCTATTA 8716
```

RESULT 4

```
ABD33504
ID   ABD33504 standard; DNA; 93544 BP.
XX
AC   ABD33504;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-098.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 676; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

  Query Match             21.7%;  Score 434.2;  DB 13;  Length 93544;
  Best Local Similarity   59.7%;  Pred. No. 5.3e-82;
  Matches  861;  Conservative    1;  Mismatches  559;  Indels   22;  Gaps    7;

Qy      404 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
            |||||||| || || || |||||  | |||||      ||||  |||||||  |||  ||||
Db    18099 TTTTTTTTTTTTTTTTTTGAGAGATGGAGTCTTAGCCTGTTGCCCAGGCTGGAGTGCAGT 18158

Qy      464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
            ||   |   |  | ||||||||| |||| |||    || |  ||||| |||       |||||
Db    18159 GGCACCATCTTGGGTCACTGCAACCTCCACCATGTGGGTTCAAGAGATTCTACTGCCTCA 18218

Qy      524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
             |||||||||||||| | |||   || ||  ||||||| ||| |||||||          |
Db    18219 ACCTCCCAAGTAGCTGGAATTACAGGCTTGTGCCACCACACCTTGCTAATT--TTTGTGT 18276

Qy      584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643
            ||||  ||||||| ||||  |   |  | |||||||||||   |||  ||||||||||| ||||
Db    18277 TTTATTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGATCTCAAACTCCTGGCCTCAA 18336

Qy      644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703
```

```
Db      18337 CTGATCCGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCACCATGCCT 18396

Qy        704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 763

Db      18397 GGCCACCACAGGCTTTATTTACAGGTTGCCTCATTACCTGAATTCTCTCTCTCCAAATTC 18456

Qy        764 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA---A 820

Db      18457 TTGTTACCCAAACACAGGAACATGAGATGGAGTCTCGCTCCGTCGCCCAGTCTGGAGTGC 18516

Qy        821 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 880

Db      18517 AGTGATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTTCTGCCTCAGC 18576

Qy        881 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 940

Db      18577 CTCCCAAGTAGCTGGGACTACAGGCATGCGCCACCCCACCTGGCTAA-TTTTTGTATTTT 18635

Qy        941 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1000

Db      18636 TAGTAGAGACGGGGTTTCACCATGCTGGCCAGGCTGGTCTCGAACTCCTGACCTC--GTG 18693

Qy       1001 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1060

Db      18694 ATCCACCCGCCTCAGCTTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATACCTGGC 18753

Qy       1061 CGATTACTGTCTATTTT------CTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1114

Db      18754 CAGGAACTAGACAGTTTTAGAGACATCTTTCTGAAATAGGGCTGGTGCCCCTCATTATTC 18813

Qy       1115 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1174

Db      18814 ATTAATGCTATTTTCTCCAAAACCCGGGAACCAAATAGATGTGATAGCATGGCATCCCTC 18873

Qy       1175 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1234

Db      18874 ATGATCTGAAGTTACTGTCCGTAACTGCGGTCCCATTCGATTTGGATATTGGGGGATGCC 18933

Qy       1235 TTTTCACTTATCA-CCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTG 1293

Db      18934 TGCACGTGTCTGTGTCATGTGCTGCATGTCTGCGGTGCCATCAGTCTGTGAGCTCCCTCT 18993

Qy       1294 TTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCT 1353

Db      18994 GGGCCTGCGTCCGTTTTTCTCATCCTTTGTATTATCTCAGCACTTTGTACAATACCTGGC 19053

Qy       1354 CCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCC 1413

Db      19054 ACTCAATGAATTAAAAATAATCTGAAGCGGGGGTTGGGGGGAGCTTGTCAATATAGCCCC 19113

Qy       1414 AGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAA 1473

Db      19114 ACCTTGTTTTGCCCTTGGACTTGATAAAGCCGTTTGTTGCTATGCATATGTCCATCTCTT 19173

Qy       1474 GACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCA------- 1526

Db      19174 TTCTAGACAATGAGTTACTTGGATGAATAGGCAGATCTCTTGGTTAAACCATAAAGACAG 19233

Qy       1527 GTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGAT 1586

Db      19234 GCTGGGCGCGGTGGCTCACATCTGTAATCCCAGCACTTTGAGAGGCCGAGGCGGGCAGAT 19293

Qy       1587 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1646

Db      19294 CACTTGAGGTTAGAAGTTCGATACCAGCCTGGCCAACATGGTGAAACCCAATCTCTACTA 19353

Qy       1647 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1706

Db      19354 AAAATACAAAAATTAGCCGAGTGCGGTGGCGCATGCCTGTAATCCTAGCTACTTGGGAGG 19413

Qy       1707 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1766

Db      19414 CTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGAAGCTTTCAGTGAGCCGAGATTGTGC 19473

Qy       1767 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 1826

Db      19474 CATTACACTCCAGCCTGGGCAACAGAGCGAGACTGTGTCTCAAAATGAAAACCCATAAAG 19533

Qy       1827 ATA 1829

Db      19534 ACA 19536
```

RESULT 5
ACN43914/c
ID ACN43914 standard; DNA; 160482 BP.
XX
AC ACN43914;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG26773.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 100; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 160482 BP; 44060 A; 32143 C; 33530 G; 49875 T; 0 U; 874 Other;

  Query Match   21.7%; Score 434.2; DB 11; Length 160482;
  Best Local Similarity   60.2%; Pred. No. 6e-82;
  Matches 888; Conservative   1; Mismatches 525; Indels   61; Gaps   8;

```
Qy        406 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 465

Db       5181 TGTTTTTTTTTTTCTCTCGATACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 5122

Qy        466 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 525

Db       5121 CATGATCTCGGCTCACTGAAACCTCTGCCTCCTGGGTTCACGTGATTCTCCTGCCTCAGA 5062
```

```
Qy       526 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGG--CTAATTAAAAAAATT 583
Db      5061 CTCCCAAGTAGCTGGGATTACAGGCGTGAGCCACCATGCCTGGCCCTAATGTTTTGTATT 5002

Qy       584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643
Db      5001 TTTTAGTAGAGAC-AGGTTTCGCTACATTGGTCAGGCTAGTCTCAAACTCCTGACCTCAA 4943

Qy       644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703
Db      4942 GTGATCCGCCTGCCTTGATGTACCAAAGTGCTGGGATTACAGATGTGAGCCACCCCGCCC 4883

Qy       704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 763
Db      4882 GGCCAGTCATTCACTTTCTTTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCATTCTGTC 4823

Qy       764 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 823
Db      4822 ATCCAGGCTGGAGTGT----------------------------------AGTGGCAC 4799

Qy       824 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 883
Db      4798 GATCTCGGCTCACTTCAACTTTCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTC 4739

Qy       884 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAG 943
Db      4738 TTGAGTAGCTGGGATTACAGGCATGCACCACCGCACCCGGCTAATTTTTTATATTTTTGG 4679

Qy       944 CAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATC 1003
Db      4678 GAGAGATGGGGTTTCACCATGTTAGACAGGCTGGTCTCAAACTCCTGACCTCAAGCAATC 4619

Qy      1004 TGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGA 1063
Db      4618 TGCTCGCCTTGGCCTCCTAAAGTGCTGGGATTACAGGCGTGAGCCACCACACCCGGCCAA 4559

Qy      1064 TTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGT 1123
Db      4558 --ATCATTCACTTTCTACAACAAGTTCTGTCTGTATGCCAATGAATCTCAGGACTGTACT 4501

Qy      1124 AGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTT 1183
Db      4500 TCTAGCTCTAACCTCTAAACATATATTTAAGTGTCCTTTAAAGCCAGGACTTAGATTCCT 4441

Qy      1184 TTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTT 1243
Db      4440 CAGAATCTCAAATTCAACAAATCCAAAATTCAACATATATTTTCCCTAAATTCTGCCCCT 4381

Qy      1244 ATCACCTTGTTA------TGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTT 1297
Db      4380 TTCTCTGTATTTCCTAGTTTGCTTAATAGTATCACCTTAATTCAAGGAAGAAATTTCAAA 4321

Qy      1298 TTCATTGTTACTATTTTAGTCACTGAAT--AATATGGCTTAATTTGCTTATACATCCTCC 1355
Db      4320 GTCATCTTCAATTTTTCTCTCTCTCAACGCAATAAAGCACTAATTCCTAGAAGCTCAACC 4261

Qy      1356 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1415
Db      4260 TTAAACAGTCCTCTGGAATGAGATGAAGGTTAACAGTTACTGAACACAGACCAAAACTAG 4201

Qy      1416 AGAAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGA 1475
Db      4200 GCACAGCAGCAAGGGTTGTTCAATTGAATTCAATTCTCTCAGAAGCCCTAAGGAATAGCT 4141

Qy      1476 CAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAAT------AACTCAGTT 1529
Db      4140 CCAATTTTATAGAGGAAGAAACTGAGGTTCCAAGAAGTAAGAAATTCACACAAGGAGGCT 4081

Qy      1530 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 1589
Db      4080 GGGTGTGGTGGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCCAAGGCGGGCAGATCAC 4021

Qy      1590 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 1649
Db      4020 CTGAGGTCAAGAGCTCGAGACCAGCCTGACAAACATGGTGAAACCCCATCTCTACTAAAA 3961

Qy      1650 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 1709
Db      3960 ATACAAAAATTACCCAGGTGTGGTGGTGCACGCCTGTAATCCCAGCTATTTGGGAGGCTG 3901

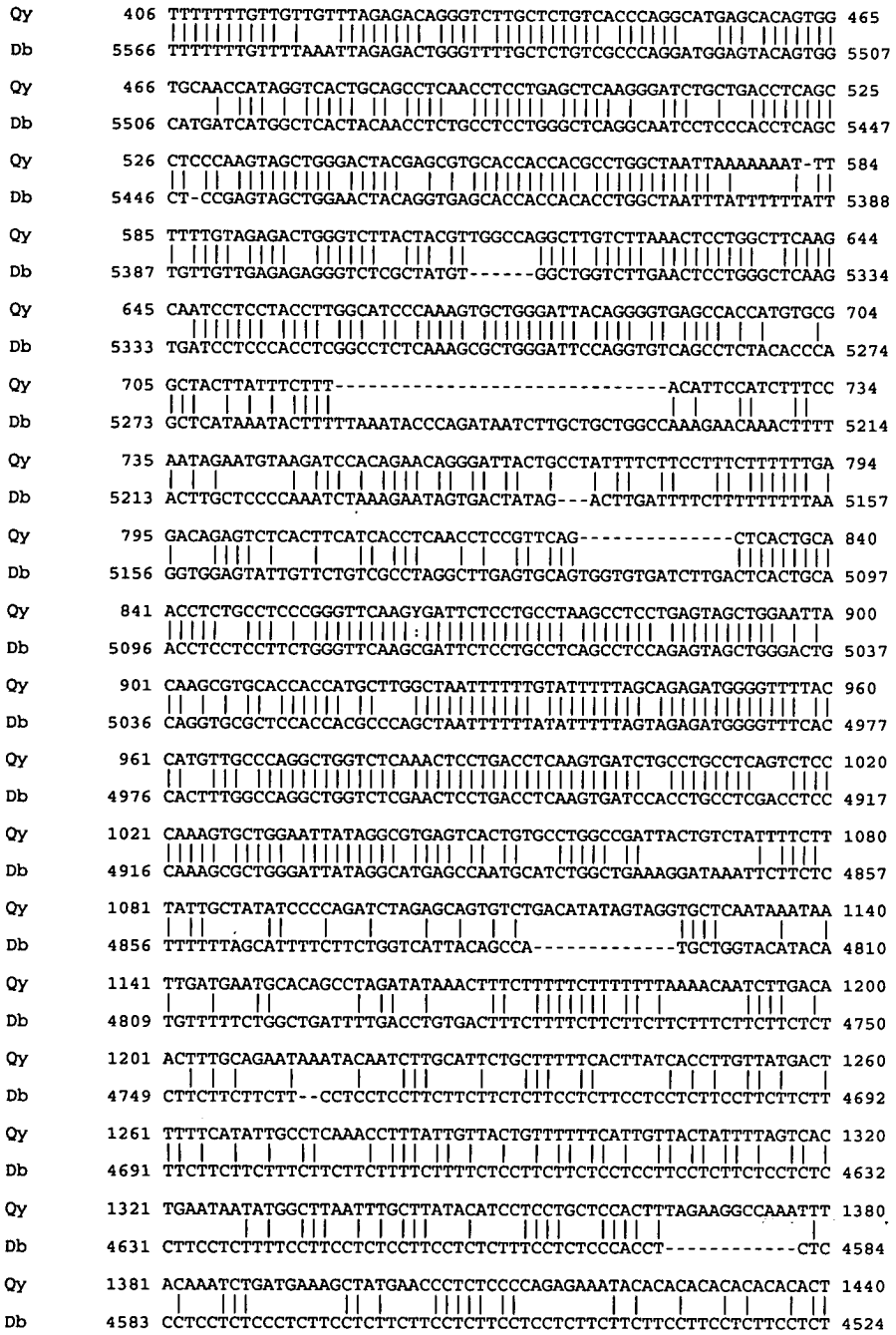
Qy      1710 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCC------GAGATCC 1763
Db      3900 ACGCAGGAGAATTGTTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATGGAGATCG 3841

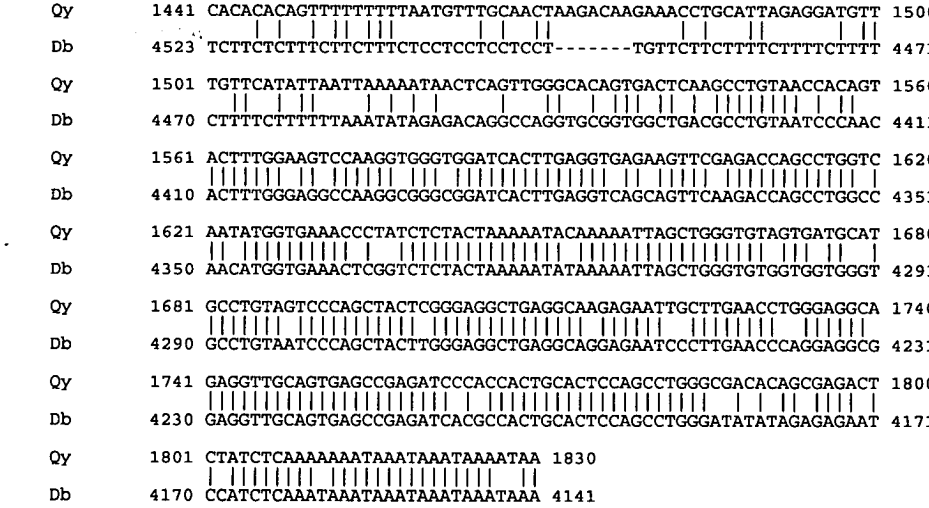
Qy      1764 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 1823
Db      3840 TGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCTGTCTCCAAAAAAAAAAAAAAA 3781

Qy      1824 AAAATAAAGGATCGGAGAGAAACAAAACTAATAAG 1858
Db      3780 AGAAAGAAAGAAAGAAAAAAATAAATTCACACAAG 3746
```

RESULT 6
ACN44410/c
ID ACN44410 standard; DNA; 91760 BP.
XX
AC ACN44410;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG41365.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 844; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 91760 BP; 21177 A; 23589 C; 24129 G; 22845 T; 0 U; 20 Other;

Query Match 21.7%; Score 433.2; DB 11; Length 91760;
Best Local Similarity 61.6%; Pred. No. 8.6e-82;
Matches 906; Conservative 1; Mismatches 474; Indels 89; Gaps 10;





RESULT 7
ABD33357/c
ID ABD33357 standard; DNA; 38538 BP.
XX
AC ABD33357;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-065.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 448; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the

```
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 38538 BP; 9418 A; 9893 C; 9523 G; 9684 T; 0 U; 20 Other;

  Query Match             21.3%;  Score 425.4;  DB 13;  Length 38538;
  Best Local Similarity   60.0%;  Pred. No. 3.2e-80;
  Matches  879;  Conservative    1;  Mismatches  552;  Indels   33;  Gaps    9;

Qy        367 CCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAG 426
Db      35304 CAAAAAATAAAAATTAAATATTTTTTGTTTCTTAGATTATAGTTAAGGGGCTTTTTTGAG 35245

Qy        427 ACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCA 485
Db      35244 ATGAAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAATGGTGTGATCTTGGCTCACAGCA 35185

Qy        486 GCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTA 545
Db      35184 ACCTCCACCTCCAGGGTTCAAGCGATTCTCCTCTCTCAGCCTCCTGAGTAGCTGGGATTA 35125

Qy        546 CGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTA 605
Db      35124 CAGGCATGCGCCACCACACCCAGCTAATTTTGTA---TTTTTAGTAGAGATGGGGTTTCT 35068

Qy        606 CTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATC 665
Db      35067 CCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCACCCTCCTTGGCTTC 35008

Qy        666 CCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC 725
Db      35007 CCGAAGTGTTGGGATTACAGGCATGAGCCACTGCGCCTGGC------CTCAGTTAAGGGA 34954

Qy        726 CATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTT 785
Db      34953 CTTTTTACATATTTATTTATTTATTTATTTATTATTATTTTGAGACGAAGTCTCGCTGTT 34894

Qy        786 CTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTC 845
Db      34893 GTTCTCCAGGCTGGAGTGCAATGGCACGACC-----------TTGGCTCACTGCAACCTC 34845

Qy        846 TGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGC 905
Db      34844 CGCCTCCCGGGTTCAAGCGATTCTCCTGTCTTGGCCTCCCAAGTAGCTGGGATTACAGGC 34785

Qy        906 GTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGT 965
Db      34784 GCCTGCCATCATGCCTGGCTAA-TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATTA 34726

Qy        966 TGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAG 1025
Db      34725 TGGCCAGGCTGGTCTTGAAGTCCTGACTTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAG 34666

Qy       1026 TGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTG 1085
Db      34665 TGCTGGGATGACAGGTGTGAGCCACCACACTCGGCCTTTATTTATTATTTTCGAGACAGG 34606

Qy       1086 CTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGAT 1145
Db      34605 GTATCTC------TCTGTTGCCCAGGCTGGAGTGCAGTGGTGCAGTCACGGCTGGTTGCA 34552

Qy       1146 GAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTT 1205
Db      34551 GCCTCAAACTCCTGGGCTCCAGCA--ATTCTCCTGCCTCAGCCTCCTGGGTAGCTGGAAC 34494

Qy       1206 GCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTC 1265
Db      34493 TACAGGTGCAAATTATCATGCCTGGCTAATTTTTAAATTTTTTGTCAAGATGGGGTCTCT 34434

Qy       1266 ATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAAT 1325
Db      34433 CTATATTGCCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATCTCCCTTCTGGGCTTCA 34374

Qy       1326 AATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAA 1385
Db      34373 AAAAGTGCTGGGATTATAAGTATGAGCCACCACCTCCGGCCTTCCCTCCTTCCTTCCTTC 34314

Qy       1386 TCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACAC 1445
Db      34313 CTTTCTGCCTTCCTTCCTTCCTTTCTCAAAAAAAAAAAAAAAAGAAGTCAGTGTTGTCTGA 34254

Qy       1446 ACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTC 1505
Db      34253 CTGGTATTGTCTGATTACCTGCATATATTGATATATTATTCCAATAAAGATGCTTGCTTT 34194

Qy       1506 ATA--TTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 1563
Db      34193 TTCCTTTGGTTAAAAAGTATCCGGCCAGGCGCGATGGCTCACGCTTGTAACCCCAGTACT 34134

Qy       1564 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 1623
Db      34133 TTGGGAGGCTGTGGTGGGCGGATCACCTGAGATCAGGAGTACGAGACCAGCCTGGCCAAC 34074

Qy       1624 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 1683
Db      34073 ATGGTAAAACCCCATCTCTACTAGAAATACAAAAATTAGCTGAGTGTAGTGGCACGCACC 34014

Qy       1684 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 1743
Db      34013 TGTAATCCCAGCTACTTGGGAGGCCAAGGGAGAAGAATTGTTTGAACCCGGGAGGCAGAG 33954

Qy       1744 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 1803
Db      33953 GTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCAGGGCGACA-AGAGAGACTCTG 33895

Qy       1804 TCTCAAAAAAATAAATAAATAAAAT 1828
Db      33894 TCTCAAAAAAAAAAAAAAAAAGCAT 33870
```

```
RESULT 8
ABK83569
ID   ABK83569 standard; cDNA; 122888 BP.
XX
AC   ABK83569;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #140.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
```

XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 140; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 122888 BP; 28761 A; 33410 C; 31919 G; 28798 T; 0 U; 0 Other;

Query Match 20.9%; Score 417.2; DB 6; Length 122888;
Best Local Similarity 61.1%; Pred. No. 2.3e-78;
Matches 894; Conservative 0; Mismatches 503; Indels 65; Gaps 11;

```
Qy      404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
Db    70248 TTTCTTTTTTATTTTTTAACACACAGGTCCTTGCTTTGTCACCCAGGCTGGAGTGCAGT 70307

Qy      464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
Db    70308 GGTGCAAACACAGCTGACAGCAACCTCAAAC-CCTGGGCTCAAGCAATCCTTCCACCTCA 70366

Qy      524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAA---------- 572
Db    70367 GCTTCCCAAGAAGCTGGGACTACAGGCATGTGACACCATGCCCAGCTAATTTTTTTTTTT 70426

Qy      573 -----TTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTT 627
Db    70427 TTTACTTTTAAAAATTTTTTGTAGAAATGGGAGTCTCACTATATTGCTTAGGCTGGTCTC 70486

Qy      628 AAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGG 687
Db    70487 AAATTCTTGGGCTTAAGCAATCCTCCTGCCTCGGCCTCCCAAAGTTGTGGGATTATAGGT 70546

Qy      688 GTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 747
```

```
Db    70547 GTG----ACCACTGCAGCCAGCCTTAATTTCTAATATGGTAAGTATTAATAGCATTTTGG 70602

Qy      748 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 807
Db    70603 GGATGCTCAATAAT-----TTTCTTCTTCTTTTTTTTTTTTTTTGAGATGGGGTCTAGC 70657

Qy      808 TTCATCACCTCAACCTCCG----------TTCAGCTCACTGCAACCTCTGCCTCCC 853
Db    70658 ACTGTCACCTAGACTGGAGTGCAATGGCGCGATTTCAGTTCACTGCAACCTCTGCCTCCT 70717

Qy      854 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 913
Db    70718 GGGTTCAAGAGATTCTCGTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCACCCGCCA 70777

Qy      914 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 973
Db    70778 CCACATCCGGCTAATTTTTTGTATTTTTAGTAGAGATAGGGTTTCGCTATGTTGGCCAGG 70837

Qy      974 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1033
Db    70838 CTGGTCTCAAACTCCTGACCTC--GTGATCCGACTGCCTCAGCCTCCCAAAGTGCTGGGA 70895

Qy     1034 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1093
Db    70896 TTACAGGTGTGAGCCACCACACTCGG----------CAATAATTTTTTAAGGTATAAAA 70944

Qy     1094 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1153
Db    70945 ATGTCCAGAGACCAAAAAGGTCAAGAACCACTGCTTTAAATACATAAATGAT-----ATC 70999

Qy     1154 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1213
Db    71000 AAACTATTTATATTATTTCAAAATTACTTTTAAAAAAATTCAACACAGAGCATAAG-CTC 71058

Qy     1214 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1273
Db    71059 CATGAAGCAGAGGTTTTTCTTTTTGTTTTATTCATTATGGAAACTAAGTCCCTAGAACAA 71118

Qy     1274 TCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGC 1333
Db    71119 TGCCTGGAACATAGTGGGTGTTCAATAAATACTGATTGAATGAATAAAGAATTATATTTT 71178

Qy     1334 TTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGA 1393
Db    71179 GGGTTTTACCCATTCTGATACGTGGACCTCAAACTCATTCTGCTGTTCAGTATTCTATTA 71238

Qy     1394 AAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTT 1453
Db    71239 TTCCTGAAGATGAAAATATCACAATAACAATTCTATTAATGAACATTTGTTCATAATTTT 71298

Qy     1454 TTTTTAA----TGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1509
Db    71299 TCATGGACATTTTTATACACAGGTAAGAGTTTCTCCAGGGTATATACCTAAGAATGAAAC 71358

Qy     1510 TAATTAAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1569
Db    71359 TGTAGTAAATAAATCTGGCCAGGCACGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA 71418

Qy     1570 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1629
Db    71419 GGCCGAGGCAAGTGGATCAC--AAGGTCAGGGGATCAAGACCATCCTGGCCAACATGGTG 71476

Qy     1630 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1689
Db    71477 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGGGGCGTGTGCCTGTAGT 71536

Qy     1690 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1749
Db    71537 CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGTTTGAACCCAGGAGGTGGAGGTTGCA 71596

Qy     1750 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 1809
Db    71597 GTGAGCCAAGATCGCACCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAC 71656
```

```
Qy       1810 AAAAATAAATAAATAAAATAAA 1831
Db      71657 AAACCAAAAAGAAAAAAAAAAA 71678


RESULT 9
AAF97850/c
ID   AAF97850 standard; DNA; 6405 BP.
XX
AC   AAF97850;
XX
DT   31-MAY-2001  (first entry)
XX
DE   Human neuroblastoma cell line NB-1 1p36 nucleotide sequence SEQ ID NO:64.
XX
KW   Human; chromosome 1; 1p36; neuroblastoma cell line; NB-1; anticancer;
KW   tumour suppressor; human 1p36 homozygosity deletion domain; tumour;
KW   diagnosis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200116311-A1.
XX
PD   08-MAR-2001.
XX
PF   31-AUG-2000; 2000WO-JP005930.
XX
PR   31-AUG-1999;   99JP-00245962.
PR   09-MAY-2000; 2000JP-00136266.
XX
PA   (HISM ) HISAMITSU PHARM CO LTD.
PA   (CHIB-) CHIBA PREFECTURE.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-226686/23.
XX
PT   Human 1p36 homozygosity deletion domain from the 36-position of first
PT   chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT   in gene diagnosis of tumors as well as in developing anti-cancer drugs.
XX
PS   Example 8; Page 93-95; 226pp; Japanese.
XX
CC   The present invention describes a homozygosity deletion domain co-
CC   existing in the 36-position of the first chromosome short arm (1p36) in
CC   human neuroblastoma. Also described are base sequences from the 1p36
CC   position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC   which are tumour suppressor genes in human neuroblastoma. The genes are
CC   tumour suppressor genes, base sequence data of which are applicable as
CC   tumour markers and reagents in studying mechanism of tumour body
CC   formation, and gene diagnosis of tumours as well as in developing anti-
CC   cancer drugs. AAF97787 to AAF97829 represent PCR primers used in the
CC   exemplification of the present invention, and AAF97830 to AAF97874
CC   represent sequences given in the exemplification of the present invention
XX
SQ   Sequence 6405 BP; 1650 A; 1449 C; 1383 G; 1923 T; 0 U; 0 Other;

  Query Match             20.8%;  Score 416.2;  DB 5;  Length 6405;
  Best Local Similarity   59.9%;  Pred. No. 1.9e-78;
  Matches  856;  Conservative    1;  Mismatches  504;  Indels   68;  Gaps    7;

Qy        403 GTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAG 462
Db       3636 GTCTTTATTTTTTATTTTTTTGAGGTGAGTTTCACTCTTGTTGCCCAGGCTGGAGTGCAA 3577

Qy        463 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 522
Db       3576 TGGCACGATCTCAGCTTACAACAACCTCTGCCTCCTGAGTTCAAGTGATTCTCCTGCCTC 3517

Qy        523 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 582
Db       3516 AGCCTCCCGAGTAGCTGGGATTACAGGCATGCACCACCACACTCAGCTGATTTTTTTGTA 3457

Qy        583 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 642
Db       3456 TTTTTAGTAGAGA-TGGGGTTCTCCATGTTGGTCAGGCTAGTCTCAAACTCCCGACCTCA 3398

Qy        643 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 702
Db       3397 GGTGATCCGCCTGCCTCGGCCTCCCGAAGTGCTGGGATTACAGGCGTGAGCCATAAGGCC 3338

Qy        703 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 762
Db       3337 TGGCCTATTTTTTACT---------------------------------------------- 3322

Qy        763 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 822
Db       3321 -TTTAAGATGGAGTTTTGCTCTTGTTGCCCAAGCTGGAGTGCAATGGCATGATC------ 3269

Qy        823 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 882
Db       3268 -----TCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGTGATTCTCTTGCCTCAGTCT 3214

Qy        883 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTA 942
Db       3213 CCCAAGTAGCTGGGATTACAGGCATGCACCACCATGCCCGGCTAA--ATTTGTATTTTTA 3156

Qy        943 GCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGAT 1002
Db       3155 GTAGAGATGGGGTTTCTCCATGTTGGTTAGGCTGGTCTCAAACTCCCAACCTC-AGTGAT 3097

Qy       1003 CTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1062
Db       3096 GCACCAGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGCCG 3037

Qy       1063 ATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAG 1122
Db       3036 TACGTATCTTTATATCAGGATGTTTATAACAATTTTCTCTTAGAACATCTTTAGAATTTA 2977

Qy       1123 TAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTT 1182
Db       2976 ATACTAGGAAAAAAGGTTT-------CTCTCGACAAACTATTAGGCTTTCTTTTACCAGA 2924

Qy       1183 TTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACT 1242
Db       2923 GTGAAAATTGTACAATTTATCTTTCCCTTTTTGCATTAAATTCCATGTTCCAAATTAAAT 2864

Qy       1243 TATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCAT 1302
Db       2863 TACAATAACCACATTAGCCGAGGCAACCTATACATTTAGTTTTTCTTGTTACATAAATTT 2804

Qy       1303 TGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCA 1362
Db       2803 TTACTTATTTATATTTTAATAGCTTCACTGTATATATTTGTGGTAAGCTGCTCAAATCCT 2744

Qy       1363 CTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATA 1422
Db       2743 TTCTAGAAGTAGTAACTGGTATAAATTATAAATAAATAAAAACTTCCTGTAATAGCCTAT 2684

Qy       1423 CACACACACACACACACTC-ACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA 1481
Db       2683 TCTGCTATTTAACACTGTAAATATGTTGTGTAGTCAAAGTTCTAGTTCCTAAAAACAAAT 2624

Qy       1482 ACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGAC 1541
Db       2623 AAAACCAAACGACACAGCAAGCTTTTAAAAGTTACATAAAGATCGGCTGGGCATCGTGGC 2564

Qy       1542 TCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAA 1601
Db       2563 TCACGCGTATAATCCCAACACTTTGGGAGGCCAAGGCGGGTGGATCACTTGAGGTCAGGA 2504

Qy       1602 GTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTA 1661
Db       2503 GTTCAAGACCAGCCTGGCCAACATGATGAAACGCTGTCTCTACTAAAAATACAAAAATTA 2444
```

```
Qy       1662 GCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAAT 1721
Db       2443 GCTGGGTGTGATGGCACATGACTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT 2384

Qy       1722 TGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCC 1781
Db       2383 CACTTGAACCCGGGAGATGAAGGTTGCAGTGAGCCAAGGTCACACCACTGCACTCTAGCC 2324

Qy       1782 TGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 1830
Db       2323 TGGATGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAGTTGCACAAA 2275
```

RESULT 10
ADP13332/c
ID ADP13332 standard; DNA; 126990 BP.
XX
AC ADP13332;
XX
DT 26-AUG-2004 (first entry)
XX
DE Renal cell carcinoma differentially expressed gene #68.
XX
KW ds; diagnosis; non-blood disease; solid tumor; gene expression;
KW peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW head/neck cancer; differential expression.
XX
OS Homo sapiens.
XX
PN WO2004048933-A2.
XX
PD 10-JUN-2004.
XX
PF 21-NOV-2003; 2003WO-US037481.
XX
PR 21-NOV-2002; 2002US-0427982P.
PR 03-APR-2003; 2003US-0459782P.
XX
PA (AMHP) WYETH.
PA (TWIN/) TWINE N C.
PA (BURC/) BURCZYNSKI M E.
PA (TREP/) TREPICCHIO W L.
PA (DORN/) DORNER A.
PA (STOV/) STOVER J A.
PA (SLON/) SLONI D K.
XX
PI Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
PI Sloni DK;
XX
DR WPI; 2004-460799/43.
XX
PT Diagnosing non-blood disease such as solid tumor, involves comparing
PT differential expression profile of specific genes in peripheral blood
PT sample of subject with reference expression profile of specific genes.
XX
PS Disclosure; SEQ ID NO 68; 350pp; English.
XX
CC The invention relate to a method of diagnosing (M1) non-blood disease
CC such as solid tumor by providing peripheral blood sample of human having
CC non-blood disease, and comparing an expression profile of specific genes
CC in the peripheral blood sample to reference expression profile of the
CC genes, where each of the genes is differentially expressed in peripheral
CC blood mononuclear cells (PBMCs) of patients having the disease as
CC compared to PBMCs of normal humans. The method is useful for diagnosing
CC non-blood disease such as solid tumor. The solid tumor is chosen from
CC renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC sample is a whole blood sample (claimed). (M1) is useful for identifying
CC genes that are differentially expressed in peripheral blood samples
CC isolated at different stages of progression, development or treatment of
CC RCC and/or other solid tumors. This sequence corresponds to a gene that
CC is differentially expressed and detected by the method of the invention.
CC (Note: this sequence is not given as part of the printed specification
CC but was obtained from WIPO in electronic format at
CC ftp.wipo./pub/published_pct_sequences).
XX
SQ Sequence 126990 BP; 36683 A; 27389 C; 27065 G; 35853 T; 0 U; 0 Other;

Query Match 20.8%; Score 416; DB 12; Length 126990;
Best Local Similarity 60.2%; Pred. No. 4.2e-78;
Matches 859; Conservative 1; Mismatches 536; Indels 32; Gaps 9;

```
Qy        411 TTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAA 470
Db      27332 TTTCTCTAATGCTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGA 27273

Qy        471 CCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCC 530
Db      27272 TCTCA----------CCTCCACCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCT-CC 27225

Qy        531 AAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGT 590
Db      27224 TGGTAGCTGGGAATACAGGCGTGCACCACCATGCCCGGCTAATTTTTATA--TTTTCAGT 27167

Qy        591 AGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCC 650
Db      27166 AGAGTCGCAGTTTCACCATGTTTCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCC 27107

Qy        651 TCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACT 710
Db      27106 ACCCGCCGCAGCCTCCCAACGTGCTGGGATAATAGGGGTGAGCCACTGCACCTGGCCTTT 27047

Qy        711 TATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGC 770
Db      27046 CTGTACTCTTTATTCTATTTCACTGGTCTGTTTCTTTTGGTTTCAGTATCATAGTACTTT 26987

Qy        771 CTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA 830
Db      26986 TGTTTCTGTAGCTTAAAATAT--ATATATTTTAAGACTGGAATGCAGTGGCATGATCTCA 26929

Qy        831 GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTA 890
Db      26928 GCTCACTGCAACCTCCACCTCCCAGGCTCAAATTATTCTTGTGACTCAGCCTCCTGAGTA 26869

Qy        891 GCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGAT 950
Db      26868 GCTGGGATTACAGGCAGGCACCACCATGCCCCACTAA-TTTTTGTATTTTCAGTAGAGTG 26810

Qy        951 GGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGC 1010
Db      26809 GAAGTTTCACCATGTTTCCTAGGCTGATCTTAAACTCCTGACCTCGAGTGATCCGCCCGC 26750

Qy       1011 CTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGT 1070
Db      26749 GTCAGCCTCCCAAAGTGCTGGGATAACAGGCATAAGCTGCTGCGTGCAGCCTTTCTGTAT 26690

Qy       1071 CTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCT 1130
Db      26689 TCTTTATTCTGTTTCATTGATCTGTTCCTTTTGGTTTCAGTATCATACTATTTTTGTTTT 26630

Qy       1131 CAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAAC 1190
Db      26629 GTTTCTATAGCTTAAAAATATGTATTTT---TTTTGAGATAGGGTCTCACTGTGTCACCC 26573

Qy       1191 AATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCT 1250
Db      26572 AGGCTGGAGTGTACTGGTGTGATCATGGCTCACTGCATCCTCAACCTCCTATACTCAA-T 26514

Qy       1251 TGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTA 1310
Db      26513 TGATCCTCCCACCTCAGCCTCCCAAGTAGCTGCACCACCATGTTTGGCTAATTTTTAAAT 26454

Qy       1311 TTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAA 1370
Db      26453 TTTTTGTAGAAACAGAGTCTTGCCATGTCACCCAGGCTGGTCTTGAACTCCTGGGCTGAA 26394
```

```
Qy       1371 GGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACA 1430
Db      26393 G---AGATCCGCTGACCTCATCCTCCCAAAGTGCTGGAATTACAGGTGTGCACCACCTTG 26337

Qy       1431 CACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA--------A 1482
Db      26336 CCCAGCCAAAAATATGTTTTTGGCAGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGG 26277

Qy       1483 CCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACT 1542
Db      26276 CCAACATGGCAAAACCTTGTCTCTACTAAAGAATACAAAAATTATCTGGGCGCAGTGGCT 26217

Qy       1543 CAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAG 1602
Db      26216 CATGTCTGTAAATCCAGTACTCTGGAAGGCCAAGGCAGGTGGATCACCTGAGGTCAGGAG 26157

Qy       1603 TTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAG 1662
Db      26156 TTTAAGACCAGCCTGGCCAACCTGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAG 26097

Qy       1663 CTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATT 1722
Db      26096 CTGGGCGTGGTGACCGGTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 26037

Qy       1723 GCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCT 1782
Db      26036 ACTTGAACCTGGGAGGTAGAGGTTGCAGTGAGCCGAGATCATGCCATTGCACTCTAGCCT 25977

Qy       1783 GGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 1830
Db      25976 GGGTGACAGAGCGAGATTCTGTCTCAAAAAAGAAAAAAAGAAAAAAA 25929
```

RESULT 11
ACN44170/c
ID ACN44170 standard; DNA; 196686 BP.
XX
AC ACN44170;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG39530.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 484; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 196686 BP; 53978 A; 42758 C; 43862 G; 55372 T; 0 U; 716 Other;

Query Match 20.7%; Score 414.2; DB 11; Length 196686;
Best Local Similarity 60.7%; Pred. No. 1.1e-77;
Matches 869; Conservative 1; Mismatches 464; Indels 97; Gaps 8;

```
Qy        405 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 464
Db      98993 TTCATTCATTTATTCATTCGAGATGGGGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAATG 98934

Qy        465 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 524
Db      98933 GTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGGATTCAAGTGATTCTCCTGCCTCAG 98874

Qy        525 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 584
Db      98873 CCTCCTGAGTAACTGGGATTACAGACACCTGCCACCACACCCGGCTAATTTTTGTA--TT 98816

Qy        585 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 644
Db      98815 TTTAGTAAAGATGGAGTTTCACCATGTTGGCCAGCCTGGTCTCAAACTCCTGACCTCAAA 98756

Qy        645 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 704
Db      98755 GGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCA 98696

Qy        705 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 764
Db      98695 GCCATGTACATTTATAATTGTACTATATCTCCTTTGGT--------CACCATCTTGTCTT 98644

Qy        765 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 824
Db      98643 TTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCGCCCAGGCTGG 98584

Qy        825 CGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTC 870
Db      98583 AGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCTGGCTCACGCCATTCTC 98524

Qy        871 CTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTT 930
Db      98523 CTGCCTCAGCCTCCAGAGTAGCTGGGACTACATGTGCCCACCACCACGCCCGGAGAATTT 98464

Qy        931 TTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTG 990
Db      98463 TTTGTATTTTTAATGGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTG 98404

Qy        991 ACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCAC 1050
Db      98403 ACCTC--GTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGAGTGAGCCAC 98346

Qy       1051 TGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTG 1110
Db      98345 TGGGCCCGGCCACCACCTCGTCTTCTGATTGCTTTTTTTTTTCACTTCTGGAGAG----- 98291

Qy       1111 TCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTT 1170
Db      98290 ----------------------------------------------------AGGAAATC 98283

Qy       1171 TCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTC 1230
Db      98282 TATATGTTCTTCACCATATCTCTAATACCTTCTTTACCTTCTCTCTAATATCTAATACTA 98223

Qy       1231 TGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTA 1290
Db      98222 TATATACATACATATAGTGTGTATATATATATGCATGTATATATAGTATTATATGTATAT 98163
```

```
Qy       1291 CTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACAT 1350
Db      98162 ATATAATAACCAGGTTCTTATTTATGTTATCCATTGCTTGAATTAACAATGTTTTTTCA- 98104

Qy       1351 CCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTC 1410
Db      98103 -------AAACCTATAAAAAGACCTTAAACAAGGTAAGGATGAGAGGAACCTTGAGTTGC 98051

Qy       1411 CCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAAC 1470
Db      98050 CCAAGATAAAATTTTGTCTCAAAGTCTCTCACTTTGTTTCATGTGTTTAATCTCTCAATA 97991

Qy       1471 TAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAA----CTCA 1526
Db      97990 TATTTAATCTATTGGCAATTACAATAATTTTTGTTTTGGTCACAAAAAATAAGCTCCTGG 97931

Qy       1527 GTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGAT 1586
Db      97930 GCTGGACACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGAGGGTGGAT 97871

Qy       1587 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1646
Db      97870 CAC--CAGGTCAGGAGATGGAGACCATCCTGGCCAACATGGTAAAACTCTGTCTCTACTA 97813

Qy       1647 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1706
Db      97812 CAAATACAAAAATTAGCTGGGCGTGGTGGCATGCGCCTGTAGTCCCAGCTACTCGGGAGG 97753

Qy       1707 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1766
Db      97752 CTGAGGCAGGAGAACTGCTTGAACCTGGGAGGCAGACGTTGCAGCGAGCTGAGATCACGC 97693

Qy       1767 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAA 1817
Db      97692 CACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCTATCTCAAAAAATAAA 97642
```

RESULT 12
ADP68568/c
ID ADP68568 standard; cDNA; 86000 BP.
XX
AC ADP68568;
XX
DT 09-SEP-2004 (first entry)
XX
DE Human PPAR-alpha cDNA.
XX
KW cytostatic; gene therapy; PPAR-alpha;
KW peroxisome proliferator-activated receptor-alpha; PPAR-alpha modulator;
KW PPAR-alpha associated disorder; hyperproliferative disorder; human; gene;
KW ss.
XX
OS Homo sapiens.
XX
PN US2004115637-A1.
XX
PD 17-JUN-2004.
XX
PF 11-DEC-2002; 2002US-00317500.
XX
PR 11-DEC-2002; 2002US-00317500.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Mckay R, Dobie KW;
XX
DR WPI; 2004-449378/42.
XX
PT New oligonucleotide compound that inhibits expression of PPAR-alpha,
PT useful for preparing a composition for treating hyperproliferative
PT disorders, e.g. cancer.
XX
PS Claim 1; SEQ ID NO 4; 121pp; English.
XX
CC The invention describes a compound, having a sequence comprising 8-80 bp
CC targeted to a nucleic acid encoding PPAR-alpha (peroxisome proliferator-
CC activated receptor-alpha), that specifically hybridises with the nucleic
CC acid encoding PPAR-alpha comprising 86001-bp sequence and inhibits
CC expression of PPAR-alpha. Also described are: a method of inhibiting the
CC expression of PPAR-alpha in cells or tissues; a method of screening for a
CC modulator of PPAR-alpha; a diagnostic method for identifying a disease
CC state; a kit or assay device comprising the compound; and a method of
CC treating an animal having a disease or condition associated with PPAR-
CC alpha. The oligonucleotide compound is useful for preparing a composition
CC for treating hyperproliferative disorder e.g. cancer. This sequence
CC represents a human peroxisome proliferator-activated receptor-alpha (PPAR
CC -alpha) cDNA.
XX
SQ Sequence 86000 BP; 22822 A; 19623 C; 20377 G; 23178 T; 0 U; 0 Other;

Query Match 20.7%; Score 413.6; DB 12; Length 86000;
Best Local Similarity 61.0%; Pred. No. 1.2e-77;
Matches 878; Conservative 1; Mismatches 490; Indels 71; Gaps 10;

```
Qy        405 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 464
Db      69228 TTTTTTTTTCTTTTTTTTTGAGACGGAGTCGCACTCTGTCGCCCAGGCTGGAGTGCAGTG 69169

Qy        465 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 524
Db      69168 GTGCGATCTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCTATTCTCCTGTCTCAT 69109

Qy        525 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 584
Db      69108 CCTCCCAAGTAGCTGGGATTACAGGTGTGCACTACCATGCCCAGCTAATTTTTGTA--TT 69051

Qy        585 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 644
Db      69050 TTTATTAGAGACGGGGTTTTGCCATGTTCGCCAGGCTGGTCTCAAACTCCTGACCTCAAG 68991

Qy        645 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 704
Db      68990 TGATCCACCCGCCTCGGCCT-CCAAAGTGCTGGGATTACAGGTGTGAACCACCGCACCTG 68932

Qy        705 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 764
Db      68931 GGAGACTTTTATTTTCTTTCTTTCTTTCTTTTTCTTTTTTTTTTTTTTGTGAGACGGAAT 68872

Qy        765 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 824
Db      68871 CTCCCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCAATC---------------------- 68833

Qy        825 CGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 884
Db      68832 ---TCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTTCTGCCTCAGCCTCC 68776

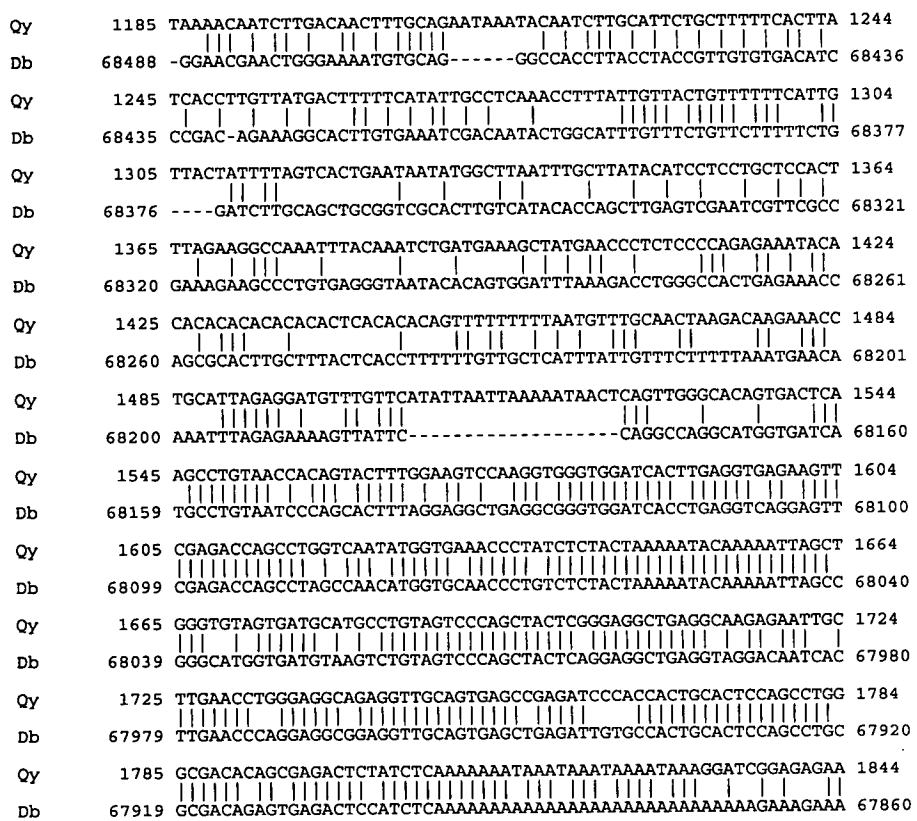
Qy        885 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 944
Db      68775 CGAGTAGCTGGGACTACAGGCATGTGCCACCACACCCAGCTAA-TTTTTGCATTCTTAGT 68717

Qy        945 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1004
Db      68716 AGAGACGGGGTTTCACCATATTGGCCAGGCTGGTCTCAAACTCCTGACCTGAGGTGATCC 68657

Qy       1005 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGAT 1064
Db      68656 ACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAACCACCGCGCCTGGGCGA- 68598

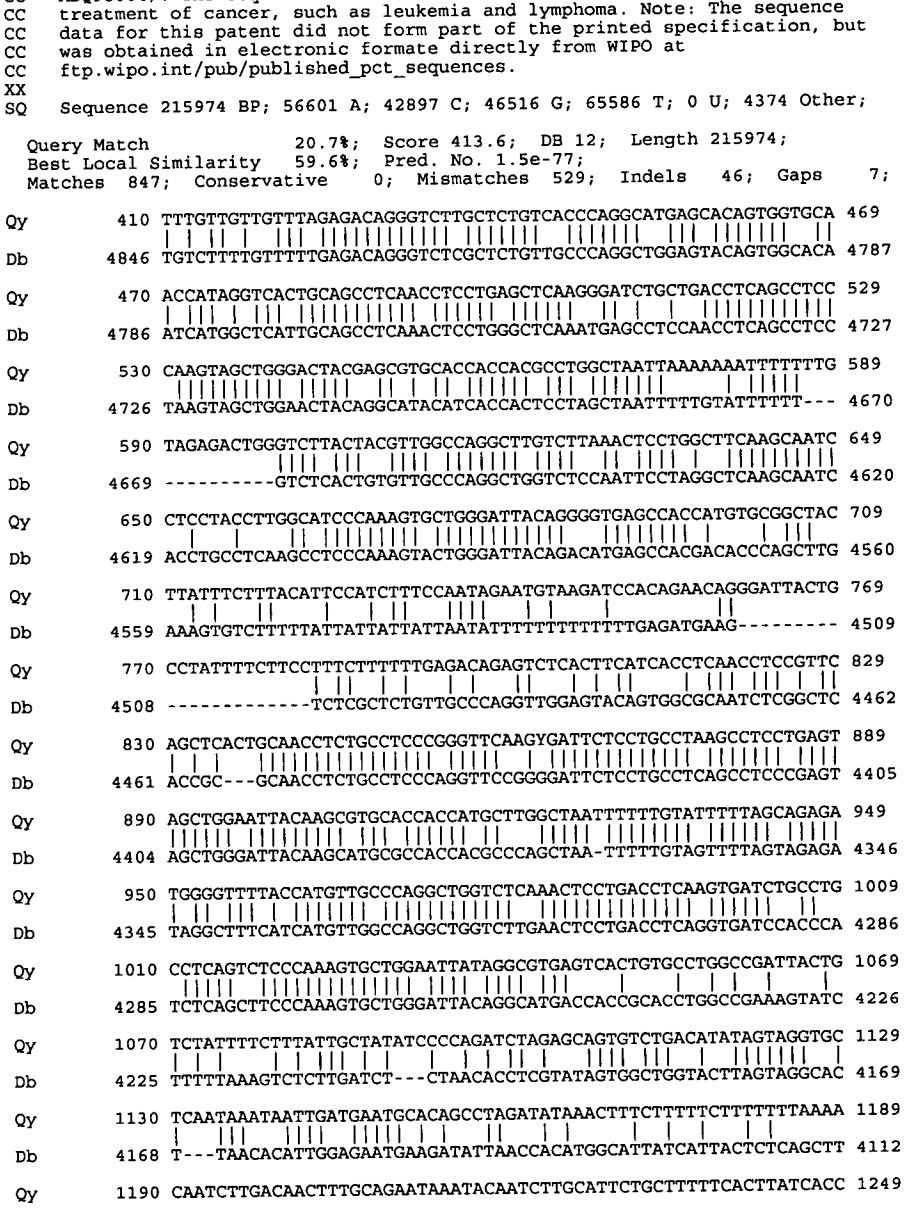
Qy       1065 TACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTA 1124
Db      68597 ----CTTTTATTTTCTTAGGCCACTCTGATTACAAAATGACTTGGGTGTGAAATGTTGAA 68542

Qy       1125 GGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTT 1184
Db      68541 GCACATTCATCCATAGCTTGATCTGTTCCCTTGGAGCAAGGAAGGGGAACTGA------- 68489
```



RESULT 13
ADQ97523/c
ID ADQ97523 standard; DNA; 215974 BP.
XX
AC ADQ97523;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD09-008, SEQ ID 500.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 500; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;

Query Match 20.7%; Score 413.6; DB 12; Length 215974;
Best Local Similarity 59.6%; Pred. No. 1.5e-77;
Matches 847; Conservative 0; Mismatches 529; Indels 46; Gaps 7;



```
Db       4111 AAATGGCTTGCCTCAAGCCACGCAGCTGGTAGGTGGCAGAGTTGGAATGCAAACACAAAT 4052

Qy       1250 TTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACT 1309

Db       4051 TTGTTTGATTCCACAAATCTTTGCCCTTTCCACAAACAAGGCCCTACATGTCCCCTTACA 3992

Qy       1310 ATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGA 1369

Db       3991 TATTTTTCAGGTGTACATTATCTTTCCAGCATTGTGAAC-TCCTAGTGAAAGCCTAGATC 3933

Qy       1370 AGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACAC 1429

Db       3932 TCACTAATAGCAATGACCTCTTCTAGCTGTTATTTATTGAGTGCTTATGATGTGCCAGGC 3873

Qy       1430 ACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCAT 1489

Db       3872 ATTGCATTTAGCACTTCACACATATGATCTCATTTAATTTTCACAACAATCCCATGAGAT 3813

Qy       1490 TAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCT 1549

Db       3812 TGGCACTATTATTATGGCCTTTAACAGAAAGGAAACAGGCCAGCATGGTGGCTCATGCCT 3753

Qy       1550 GTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGA 1609

Db       3752 ATAATCCTAGCACTTTGGGAGGCTGAGGTGGGTTGATCACCTGAGGTCAGGAGTTCAAGA 3693

Qy       1610 CCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTG 1669

Db       3692 CCAACCTGGCCAACATAGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCA 3633

Qy       1670 TAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAA 1729

Db       3632 TGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAA 3573

Qy       1730 CCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGAC 1789

Db       3572 CCCAGGAGGCGGAGGTTGCAGTGAGCAGAGATTGCACTGTTGCACTCCAGCCTGGGAGAC 3513

Qy       1790 ACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 1831

Db       3512 AGAGCAAGACTCCGTCTCAAAAAGAAACAGAGAGGAAACAGA 3471
```

```
RESULT 14
AAL42369/c
ID   AAL42369 standard; DNA; 13670 BP.
XX
AC   AAL42369;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Human Guanine nucleotide binding protein gamma 7 (GNG7) gene sequence.
XX
KW   Human; gene; ds; guanine nucleotide binding protein gamma 7; G protein;
KW   GNG7; novel polymorphic site; drug screening; gene therapy;
KW   GNG7-related disease; pancreatic cancer; GNG7 haplotyping;
KW   GNG7 genotyping.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   allele          replace(3900, T)
FT                   /*tag=  a
FT                   /note= "Single nucleotide polymorphism"
FT   allele          replace(3975, G)
FT                   /*tag=  b
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4005. .9671
FT                   /*tag=  c
FT                   /product= "Guanine nucleotide binding protein gamma 7"
FT   exon            4005. .4085
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4086. .9545
FT                   /*tag=  e
FT                   /number= 1
FT   exon            9546. .9671
FT                   /*tag=  f
FT                   /number= 2
FT   allele          replace(9958, A)
FT                   /*tag=  g
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200218647-A1.
XX
PD   07-MAR-2002.
XX
PF   23-AUG-2001; 2001WO-US026279.
XX
PR   25-AUG-2000; 2000US-0228234P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Finkel K,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-315548/35.
DR   P-PSDB; AAO14753.
XX
PT   Novel genetic variants of guanine nucleotide binding protein, gamma 7
PT   gene useful in studying expression and function of the protein, and for
PT   screening drugs to treat diseases e.g. pancreatic cancer.
XX
PS   Claim 1; Fig 1; 62pp; English.
XX
CC   The invention comprises the nucleotide and amino acid sequences of the
CC   human guanine nucleotide binding protein (G protein) gamma 7 (GNG7). The
CC   invention specifically relates to the discovery of three novel
CC   polymorphic sites in the GNG7 gene. The GNG7 nucleotide and protein
CC   sequences are useful for screening for drugs which target GNG7 and may be
CC   used to treat (gene therapy) GNG7-related diseases (e.g. pancreatic
CC   cancer). The GNG7 nucleotide sequence can be used for haplotyping GNG7,
CC   genotyping GNG7 and predicting the haplotype pair for GNG7 in an
CC   individual. The present DNA sequence represents the human GNG7 gene
CC   sequence
XX
SQ   Sequence 13670 BP; 3047 A; 3946 C; 3680 G; 2997 T; 0 U; 0 Other;

  Query Match           20.7%;  Score 413.4;  DB 6;  Length 13670;
  Best Local Similarity   60.6%;  Pred. No. 8.9e-78;
  Matches  876;  Conservative    1;  Mismatches  512;  Indels   56;  Gaps   10;

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCAT-GAGCACAG 462

Db       5900 TTTGTTTTTGTTTTTTTTCTGAGATGGAGTCTCACCCTGTCGCTTGGGGTTGGAGGGCAG 5841

Qy        463 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 522

Db       5840 TGGCGCAATCTCAGCTCACTGCAACCTCTGCCTCTCGGGTTCAAGCTATTCTCCTGACTC 5781

Qy        523 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 582

Db       5780 AGTCTCCCAAGTAGCTGGGACTATAGGAGCCTGCCACCACGCCTGGCTAATTTTTTTGTA 5721

Qy        583 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 642

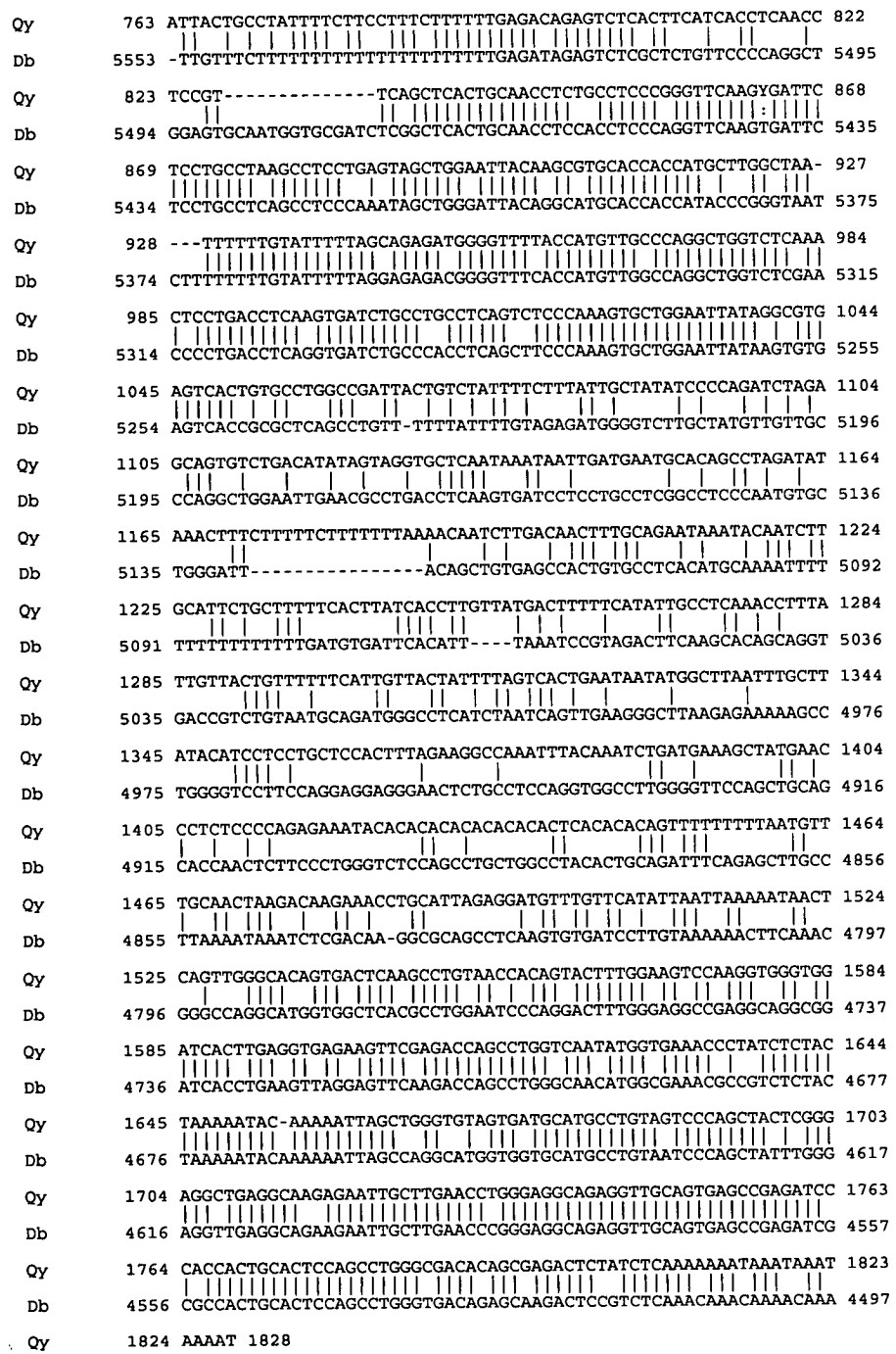
Db       5720 ATTTTAGTAGAGATGGGGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTGAC--CT 5663

Qy        643 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 702

Db       5662 TGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGCGATTCCAGGCGTGAGTCACCGTGCC 5603

Qy        703 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 762

Db       5602 TGGCCAGTTAAATGTTTTTTTAGATGTTGCTGTGAAGGTTGTTTTTTTC---------- 5554
```



```
Qy       763 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 822
Db      5553 -TTGTTTCTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTCTCGCTCTGTTCCCCAGGCT 5495

Qy       823 TCCGT-------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 868
Db      5494 GGAGTGCAATGGTGCGATCTCGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTC 5435

Qy       869 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA- 927
Db      5434 TCCTGCCTCAGCCTCCCAAATAGCTGGGATTACAGGCATGCACCACCATACCCGGGTAAT 5375

Qy       928 ---TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAA 984
Db      5374 CTTTTTTTTGTATTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA 5315

Qy       985 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTG 1044
Db      5314 CCCCTGACCTCAGGTGATCTGCCCACCTCAGCTTCCCAAAGTGCTGGAATTATAAGTGTG 5255

Qy      1045 AGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGA 1104
Db      5254 AGTCACCGCGCTCAGCCTGTT-TTTTATTTTGTAGAGATGGGGTCTTGCTATGTTGTTGC 5196

Qy      1105 GCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATAT 1164
Db      5195 CCAGGCTGGAATTGAACGCCTGACCTCAAGTGATCCTCCTGCCTCGGCCTCCCAATGTGC 5136

Qy      1165 AAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTT 1224
Db      5135 TGGGATT---------------ACAGCTGTGAGCCACTGTGCCTCACATGCAAAATTTT 5092

Qy      1225 GCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1284
Db      5091 TTTTTTTTTTTTGATGTGATTCACATT----TAAATCCGTAGACTTCAAGCACAGCAGGT 5036

Qy      1285 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1344
Db      5035 GACCGTCTGTAATGCAGATGGGCCTCATCTAATCAGTTGAAGGGCTTAAGAGAAAAAGCC 4976

Qy      1345 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1404
Db      4975 TGGGGTCCTTCCAGGAGGAGGGAACTCTGCCTCCAGGTGGCCTTGGGGTTCCAGCTGCAG 4916

Qy      1405 CCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTT 1464
Db      4915 CACCAACTCTTCCCTGGGTCTCCAGCCTGCTGGCCTACACTGCAGATTTCAGAGCTTGCC 4856

Qy      1465 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1524
Db      4855 TTAAAATAAATCTCGACAA-GGCGCAGCCTCAAGTGTGATCCTTGTAAAAAACTTCAAAC 4797

Qy      1525 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1584
Db      4796 GGGCCAGGCATGGTGGCTCACGCCTGGAATCCCAGGACTTTGGGAGGCCGAGGCAGGCGG 4737

Qy      1585 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1644
Db      4736 ATCACCTGAAGTTAGGAGTTCAAGACCAGCCTGGGCAACATGGCGAAACGCCGTCTCTAC 4677

Qy      1645 TAAAAATAC-AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGG 1703
Db      4676 TAAAAATACAAAAAATTAGCCAGGCATGGTGGTGCATGCCTGTAATCCCAGCTATTTGGG 4617

Qy      1704 AGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCC 1763
Db      4616 AGGTTGAGGCAGAAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCG 4557

Qy      1764 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 1823
Db      4556 CGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAACAAACAAAACAAA 4497

Qy      1824 AAAAT 1828
Db      4496 CAAAT 4492
```

RESULT 15
ABA16909
ID ABA16909 standard; DNA; 6519 BP.
XX
AC ABA16909;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9240.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
PS Disclosure; SEQ ID NO 9240; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6519 BP; 1661 A; 1577 C; 1542 G; 1739 T; 0 U; 0 Other;

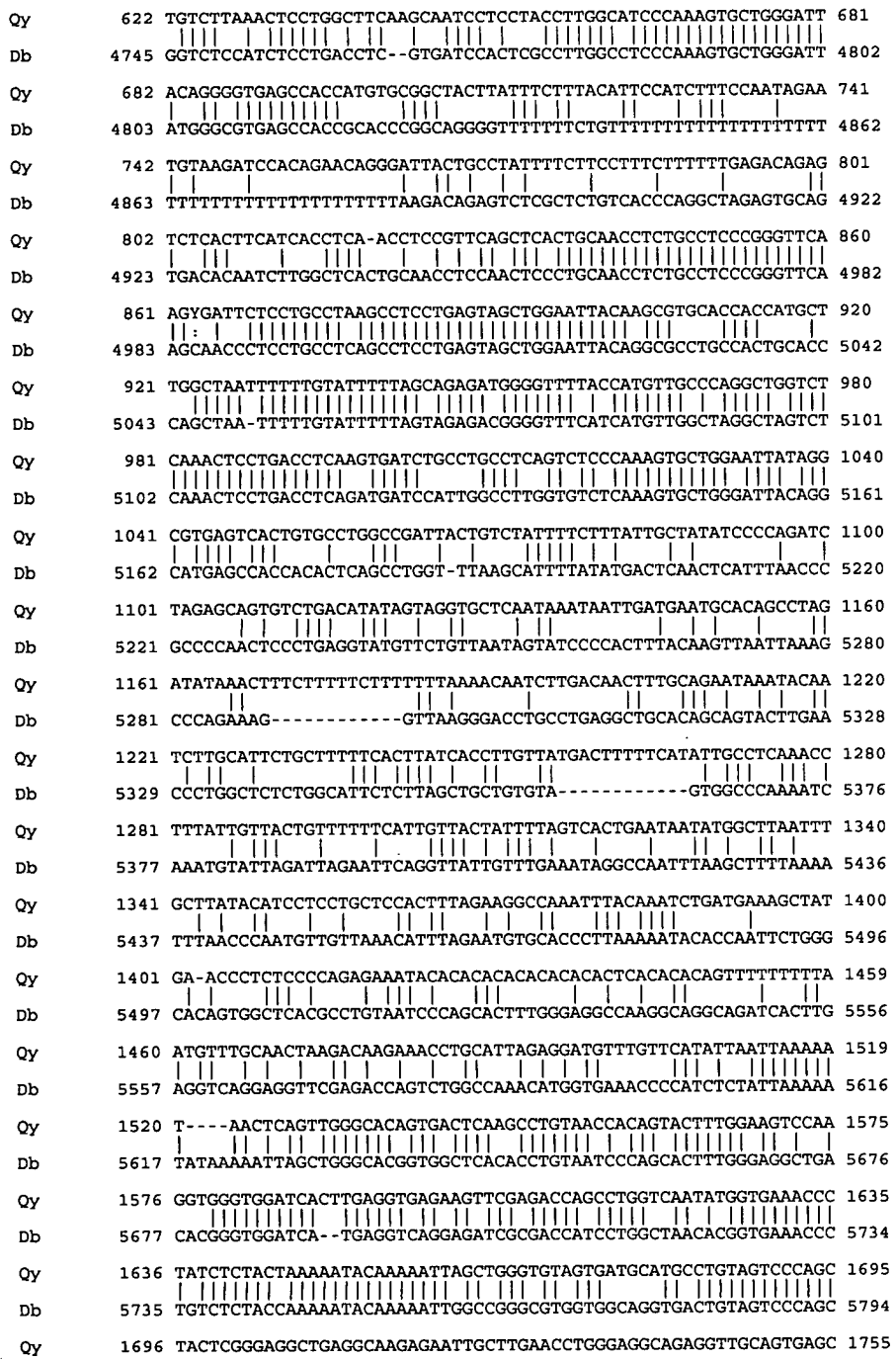
Query Match 20.7%; Score 413.2; DB 5; Length 6519;
Best Local Similarity 59.8%; Pred. No. 8.3e-78;
Matches 865; Conservative 1; Mismatches 544; Indels 36; Gaps 9;

Qy 382 TTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCT 441
Db 4505 TTTTATGTTTTATTTTTTGGGTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCGCTCT 4564

Qy 442 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 501
Db 4565 GTCACCCAGGCTGGAGTGCAGTGACGTGATCTCAGCTCACTGCAAGCTCCGTTTCCCAGG 4624

Qy 502 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 561
Db 4625 TTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGTGCCCACGACCA 4684

Qy 562 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 621
Db 4685 TGCCCGGCTAATTTTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGAT 4744



Search completed: February 9, 2006, 13:51:03
Job time : 1387.77 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2006, 13:34:47 ; Search time 9440.43 Seconds
(without alignments)
9917.025 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 1998.6
Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	505.4	25.3	507	5	BU785040	BU785040 in43g09.x
	2	344.2	17.2	4087	4	BC024593	BC024593 Homo sapi
	3	341	17.1	3990	4	HSM803026	AL713681 Homo sapi
c	4	337.2	16.9	15970	9	AQ839852	AQ839852 260L13-C5
c	5	315.8	15.8	897	6	CD557847	CD557847 AGENCOURT
	6	311	15.6	3095	4	CR858326	CR858326 Pongo pyg
c	7	307.2	15.4	5797	4	CR749233	CR749233 Homo sapi
c	8	301.2	15.1	1605	4	CR860521	CR860521 Pongo pyg
	9	301.2	15.1	1863	4	CR599842	CR599842 full-leng
	10	300.6	15.0	946	5	BQ958903	BQ958903 AGENCOURT
c	11	298	14.9	918	5	BQ706343	BQ706343 AGENCOURT
	12	293.6	14.7	736	8	CX780936	CX780936 HESC3_2_B
	13	290.6	14.5	3092	4	CR613629	CR613629 full-leng
c	14	288.2	14.4	855	5	BQ681302	BQ681302 AGENCOURT
c	15	287.6	14.4	769	5	BQ710315	BQ710315 AGENCOURT
c	16	285.2	14.3	4088	4	HSM808459	BX648311 Homo sapi
	17	283	14.2	3165	4	CR859576	CR859576 Pongo pyg
c	18	281.6	14.1	800	1	AU120942	AU120942 AU120942
	19	281.4	14.1	3259	4	CR860168	CR860168 Pongo pyg
	20	280.4	14.0	672	6	CA431692	CA431692 UI-H-DF0-
	21	280	14.0	3474	4	HSM800886	AL110229 Homo sapi
	22	278.8	13.9	658	9	AQ393450	AQ393450 CITBI-E1-
c	23	277.6	13.9	877	9	AQ739838	AQ739838 HS_5505_A
	24	277.2	13.9	1875	4	BC009270	BC009270 Homo sapi
c	25	276.6	13.8	3990	4	HSM803026	AL713681 Homo sapi
	26	276.4	13.8	4828	4	HSM802759	AL390128 Homo sapi
c	27	276.2	13.8	881	5	BU521286	BU521286 AGENCOURT
	28	274.8	13.7	3552	4	CR860263	CR860263 Pongo pyg
	29	273.8	13.7	736	6	CA427039	CA427039 UI-H-DF0-
c	30	273.4	13.7	629	5	BX509360	BX509360 DKFZp686J
c	31	273.4	13.7	2230	4	CR859082	CR859082 Pongo pyg
	32	272.4	13.6	652	6	CA427045	CA427045 UI-H-DF0-
	33	272.4	13.6	666	6	CA431783	CA431783 UI-H-DF0-
c	34	272.4	13.6	815	8	CX785622	CX785622 HESC3_49_
c	35	272.2	13.6	922	5	BU501973	BU501973 AGENCOURT
c	36	272	13.6	617	3	BI861844	BI861844 603388872
	37	271.8	13.6	644	7	CN480313	CN480313 UI-H-EU0-
	38	271.6	13.6	1042	5	BX377759	BX377759 BX377759
	39	271.6	13.6	2097	4	CR614786	CR614786 full-leng
c	40	271.4	13.6	603	1	AL707313	AL707313 DKFZp686P
c	41	271.2	13.6	5325	4	HSM804049	AL832738 Homo sapi
c	42	271.2	13.6	5785	4	HSM802309	AL137734 Homo sapi
c	43	270.6	13.5	2330	4	CR615928	CR615928 full-leng
	44	270.4	13.5	618	9	BZ609884	BZ609884 WHACH76TF
	45	270	13.5	665	5	BU633001	BU633001 UI-H-DF0-

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS BU785040 507 bp mRNA linear EST 11-OCT-2002
DEFINITION in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3', mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES Location/Qualifiers
source 1. .507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 25.3%; Score 505.4; DB 5; Length 507;
Best Local Similarity 99.8%; Pred. No. 1.9e-35;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1306 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 1365
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 448

Qy       1366 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 1425
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 388

Qy       1426 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 1485
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 328

Qy       1486 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 1545
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 268

Qy       1546 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 1605
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 208

Qy       1606 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 1665
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 148

Qy       1666 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCT 1725
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        147 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCT 88

Qy       1726 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 1785
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 28

Qy       1786 CGACACAGCGAGACTCTATCTCAAAAA 1812
             |||||||||||||||||||||||||||||
Db          27 CGACACAGCGAGACTCTATCTCAAAAA 1
```

RESULT 2
BC024593
LOCUS BC024593 4087 bp mRNA linear HTC 29-JUN-2004
DEFINITION Homo sapiens cDNA clone IMAGE:3914314, with apparent retained intron.
ACCESSION BC024593
VERSION BC024593.1 GI:22137609
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4087)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 4087)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 22 Row: m Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.

FEATURES Location/Qualifiers
source 1. .4087
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3914314"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH_MGC_71"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

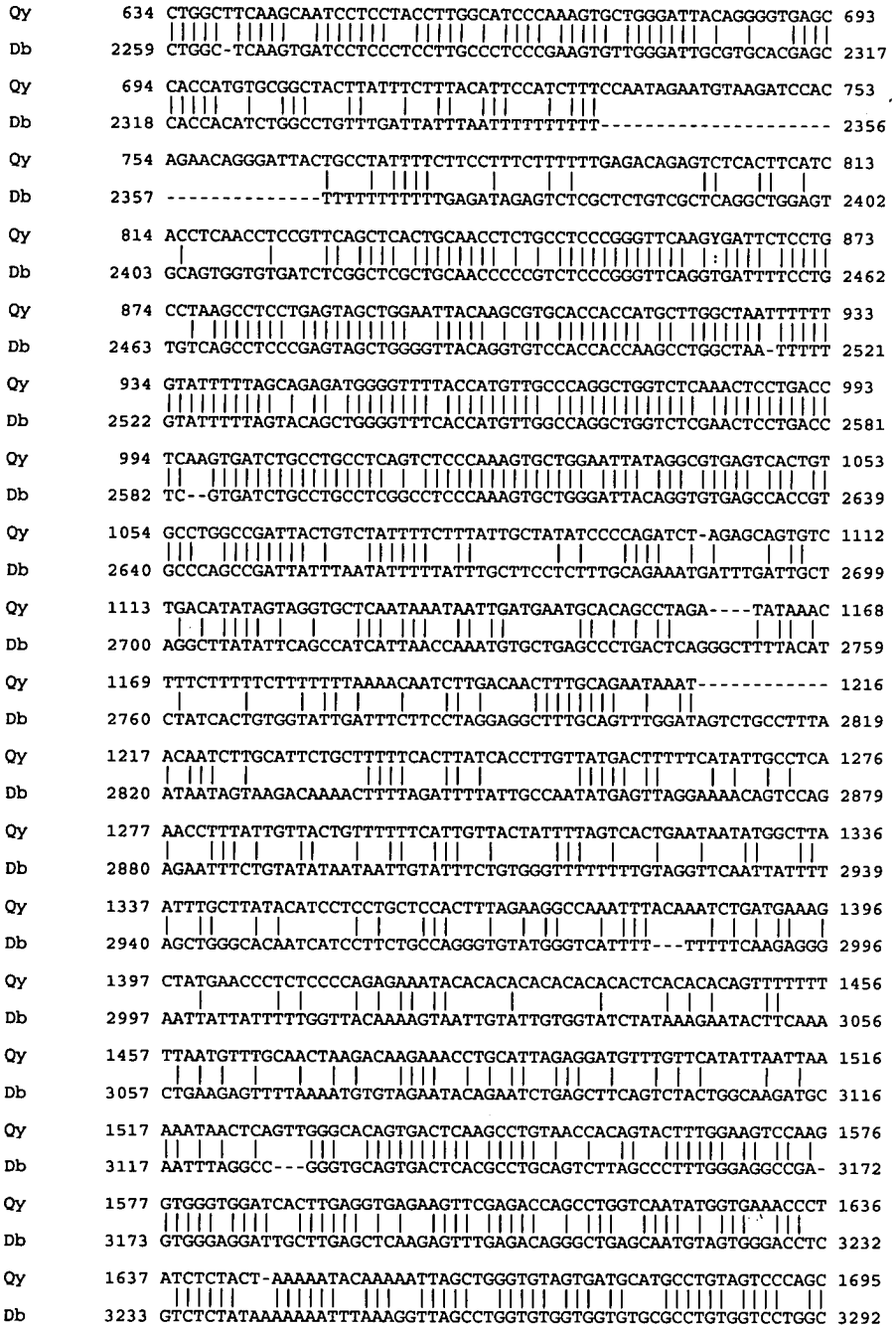
Query Match 17.2%; Score 344.2; DB 4; Length 4087;
Best Local Similarity 58.3%; Pred. No. 5.5e-22;
Matches 859; Conservative 1; Mismatches 544; Indels 69; Gaps 12;

```
Qy        399 TTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 458
              || |   ||| ||||||  ||||| ||||||||||| | |||| |||||||||  |||
Db       2019 TTTGAAATTTGTTTGTTTGTGTTTTGAGACAGGGTCTGGGTCTGCCACCCAGGCTGGAGT 2078

Qy        459 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 518
              |||||||  | ||||| ||| ||||||| |||||| | | || |||| |||   || |
Db       2079 GTAGTGGTGTGATCATAGCTCAATGCAGCCCCAACCTTCCGGGCGCAAGTGATGCTCTCA 2138

Qy        519 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 578
              ||||||||||  ||||||||||||||||  || ||  |||||||  ||    |||||  | |
Db       2139 CCTCAGCCTCCTGAGTAGCTGGGACTACAGGCATGTGCCACCACACCCTCCTAATATATA 2198

Qy        579 AAAT-----TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 633
              |     ||||||||||||   ||| | | || |||| |||| || ||||  |||||
Db       2199 TTTTTTTTATTTTTTGTAGAGATGAGGTTTCATTATGTTGCCCAGACTGGTCTGGAACTC 2258
```



```
Qy       634 CTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC 693
Db      2259 CTGGC-TCAAGTGATCCTCCCTCCTTGCCCTCCCGAAGTGTTGGGATTGCGTGCACGAGC 2317

Qy       694 CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC 753
Db      2318 CACCACATCTGGCCTGTTTGATTATTTAATTTTTTTTTT---------------------- 2356

Qy       754 AGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATC 813
Db      2357 --------------TTTTTTTTTTTGAGATAGAGTCTCGCTCTGTCGCTCAGGCTGGAGT 2402

Qy       814 ACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTG 873
Db      2403 GCAGTGGTGTGATCTCGGCTCGCTGCAACCCCCGTCTCCCGGGTTCAGGTGATTTTCCTG 2462

Qy       874 CCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT 933
Db      2463 TGTCAGCCTCCCGAGTAGCTGGGGTTACAGGTGTCCACCACCAAGCCTGGCTAA-TTTTT 2521

Qy       934 GTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACC 993
Db      2522 GTATTTTTAGTACAGCTGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACC 2581

Qy       994 TCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGT 1053
Db      2582 TC--GTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGT 2639

Qy      1054 GCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCT-AGAGCAGTGTC 1112
Db      2640 GCCCAGCCGATTATTTAATATTTTTATTTGCTTCCTCTTTGCAGAAATGATTTGATTGCT 2699

Qy      1113 TGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGA----TATAAAC 1168
Db      2700 AGGCTTATATTCAGCCATCATTAACCAAATGTGCTGAGCCCTGACTCAGGGCTTTTACAT 2759

Qy      1169 TTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAAT------------ 1216
Db      2760 CTATCACTGTGGTATTGATTTCTTCCTAGGAGGCTTTGCAGTTTGGATAGTCTGCCTTTA 2819

Qy      1217 ACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCA 1276
Db      2820 ATAATAGTAAGACAAAACTTTTAGATTTTATTGCCAATATGAGTTAGGAAAACAGTCCAG 2879

Qy      1277 AACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTA 1336
Db      2880 AGAATTTCTGTATATAATAATTGTATTTCTGTGGGTTTTTTTTGTAGGTTCAATTATTTT 2939

Qy      1337 ATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAG 1396
Db      2940 AGCTGGGCACAATCATCCTTCTGCCAGGGTGTATGGGTCATTTT---TTTTTCAAGAGGG 2996

Qy      1397 CTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTT 1456
Db      2997 AATTATTATTTTTGGTTACAAAAGTAATTGTATTGTGGTATCTATAAAGAATACTTCAAA 3056

Qy      1457 TTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAA 1516
Db      3057 CTGAAGAGTTTTAAAATGTGTAGAATACAGAATCTGAGCTTCAGTCTACTGGCAAGATGC 3116

Qy      1517 AAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAG 1576
Db      3117 AATTTAGGCC---GGGTGCAGTGACTCACGCCTGCAGTCTTAGCCCTTTGGGAGGCCGA- 3172

Qy      1577 GTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCT 1636
Db      3173 GTGGGAGGATTGCTTGAGCTCAAGAGTTTGAGACAGGGCTGAGCAATGTAGTGGGACCTC 3232

Qy      1637 ATCTCTACT-AAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGC 1695
Db      3233 GTCTCTATAAAAAAATTTAAAGGTTAGCCTGGTGTGGTGGTGTGCGCCTGTGGTCCTGGC 3292

Qy      1696 TACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGC 1755
Db      3293 TACTTGGGAGGCTGTGATGGGAGGATCGCTGGAGCGTGGGAGGTCAAGGCTGCGGTGGGT 3352

Qy      1756 CGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAAT 1815
Db      3353 CATGGTTGTGCCACTGCACTTCAGCCTGGGTGACAGAGTAAGACCCTGTCTAAAAAGCAA 3412

Qy      1816 AAATAAATAAAATAAAGGATCGGAGAGAAACAA 1848
Db      3413 AATAAGATGCAGTTTGGAGTGTAAAATAGAGAA 3445
```

```
RESULT 3
HSM803026
LOCUS       HSM803026              3990 bp    mRNA    linear   HTC 16-APR-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp761O0217 (from clone DKFZp761O0217).
ACCESSION   AL713681
VERSION     AL713681.1  GI:19584382
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3990)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp761O0217) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp761O0217
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3990
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp761O0217"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761O0217"
                     /tissue_type="amygdala"
                     /clone_lib="761 (synonym: hamy2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="cytochrome c"
     gene            1. .3990
                     /gene="DKFZp761O0217"
     CDS             61. .378
                     /gene="DKFZp761O0217"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD28485.1"
                     /db_xref="GI:19584383"
                     /db_xref="GOA:P99999"
                     /db_xref="InterPro:IPR002327"
                     /db_xref="InterPro:IPR003088"
                     /db_xref="UniProt/Swiss-Prot:P99999"
                     /translation="MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQA
                     PGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKA
                     TNE"
ORIGIN
```

```
 Query Match             17.1%;  Score 341;  DB 4;  Length 3990;
 Best Local Similarity   57.5%;  Pred. No. 1.1e-21;
 Matches  856;  Conservative    1;  Mismatches  541;  Indels   91;  Gaps   10;

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
Db       2522 TTTTCTCATTTATTTTTTCTGAGACGGAGTTTCGCCCTTTTCGCCCAGGCTGAGTACAGT 2581

Qy        464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
Db       2582 GGTGCAATC-----TCACTGCAACCTCCGCCTTCCAGTTGCAAGTGATTCTCCTGCCTCA 2636

Qy        524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
Db       2637 GCCTCCTGAGTAGCTGGGATTGCAGGCACCCGCCACCACGCCTGGCTAATTTTTGTA--T 2694

Qy        584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643
Db       2695 TTTTAGTGGAGATGGTGTTTCACCATGTTGGCCAGACTGGTCTTGGACTCCTGACCTC-- 2752

Qy        644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703
Db       2753 GTGATCCACCCACCTTGGCCTCCCAAAGTTCTAGGATTACAGGTGTGAGCCACCACGTCC 2812

Qy        704 GGCTACTTATT---------------------------------------------------- 714
Db       2813 GGCCAATTTTTCTCATTTCTATGCCTCCTATATTAAGGTCTGTGTTGGCACAGATGAGTA 2872

Qy        715 ------TCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACT 768
Db       2873 ACTGCCATGTTCTAGGTCAGTTATACCCAAGCACTTCTGGTGGTTTAAAATGTGATTCTG 2932

Qy        769 GCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC------------ 816
Db       2933 TAACTTTTTTATTTTTATTTTTTTGAGATAATTTCACTCTTGTTGCCCAGGCTGGAGTGC 2992

Qy        817 -TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCC 875
Db       2993 AATGGCGTGATCGCTGCTCACCGCAACCTCCGCCTTCCAGGTTCAAGCGATTCTCCTGAC 3052

Qy        876 TAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGT 935
Db       3053 TCAGCCTCTCAAGTAGCTGGGATTACAAGCATGCGCCACCATGCCCAGCTTATTT--TGT 3110

Qy        936 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 995
Db       3111 GTTTTTAGTAGAGACAGGGTTTCTCCATGCTGGACAGGCTGGTCTTGAACTCCCGACGTC 3170

Qy        996 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1055
Db       3171 GGGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGC 3230

Qy       1056 CTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGA 1115
Db       3231 CTGGCCAATTA---TGTAATTTTTTAAAAAGGACATTTCTATCAGGGATATATACCTTCA 3287

Qy       1116 CATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTT 1175
Db       3288 GAAATAAGGAAATAGGGGAAAAAAAGAGCACTATAAACCACATGTTTTCATTTCTAGTGC 3347

Qy       1176 TTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTT 1235
Db       3348 TTCGCTGTAAGTGGCTAGGTTGGTAGAATCAAAAACAAGGGCCAGATGTATTTAAGGGGT 3407

Qy       1236 TTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTT 1295
Db       3408 ATTCAGATGCCACCTACATGCTTATTTTGTCTAGAACAGTGCTGTCTAATAGAACTTTCT 3467

Qy       1296 TTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCC 1355
Db       3468 GTGACGATGGATATTTTGTAGACTTT------TGCTGTCCAGTGTGGTAGCCACTAACCA 3521

Qy       1356 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1415

Db       3522 CATGTGGCTGTTAAGCCCTTGAAATATAGCTAGTGTGACTAGAAAGTATTTTATTTTAAA 3581

Qy       1416 AGAAATACACACACACACACACTCACACACAGTTTTTTTTAATGTTTGCAACTAAGA 1475
Db       3582 TTTACATAGGCACAAGTGGCTAGTGGCTACTGTATTGACATTCTGGGTCTAGGACTAGAA 3641

Qy       1476 CAAGAAACCTGCATTAGAGG--ATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGC 1533
Db       3642 CCAGTGGTCTGTAACAAAAGTACTTTCTCTTTACTCTATTAAATCTAGAATTAGCCGGGC 3701

Qy       1534 ACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGA 1593
Db       3702 ATGGTCGCTCATGCCTGTGGTCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACTTGA 3761

Qy       1594 GGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC 1653
Db       3762 GGTCGGGCGTTTGAGACCAGCCTGGTCAACATGGCGAAACCCTGTCTCTACAAAAAACAT 3821

Qy       1654 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1713
Db       3822 AAAAGTTAGCCAGGTGTGGTGGTGGGCACCTGTAGTCTCAGCTACTTGGGAGGCTGAGGC 3881

Qy       1714 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1773
Db       3882 ACAAGAATCACTTGAACTTGGGAGGTGGAGGTTGCAGTGAGCCAGGATTGTGCCACTGCA 3941

Qy       1774 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAA 1822
Db       3942 CTCAAGCCTGGGTGAC-GAGTGAAACTGTCTCCAAAAAAAAAAAAAAA 3989
```

```
RESULT 4
AQ839852/c
LOCUS       AQ839852               15970 bp    DNA     linear   GSS 31-MAR-2000
DEFINITION  260L13-C56 CITB Homo sapiens genomic clone 260L13, genomic survey
            sequence.
ACCESSION   AQ839852
VERSION     AQ839852.1  GI:6652484
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 15970)
  AUTHORS   Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
            Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
            Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
            Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and
            Landes,G.M.
  TITLE     A 6-Mb high-resolution physical and transcription map encompassing
            the hereditary prostate cancer 1 (HPC1) region
  JOURNAL   Genomics 64 (1), 1-14 (2000)
   PUBMED   10708513
COMMENT     Contact: Carpten JD
            Cancer Genetics Branch
            National Human Genome Research Institute/National Institutes of
            Health
            Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
            Tel: 301 435 5626
            Fax: 301 435 5465
            Email: jdc@nhgri.nih.gov
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .15970
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="260L13"
                     /clone_lib="CITB"
ORIGIN
```

```
Query Match           16.9%;  Score 337.2;  DB 9;  Length 15970;
Best Local Similarity  57.8%;  Pred. No. 9.4e-22;
Matches  905;  Conservative    1;  Mismatches  529;  Indels  131;  Gaps   12;

Qy        398 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 457
              |||   ||||| ||| || || ||| |||    | || | ||| ||||||||||||  |||
Db       6787 TTTTTATTTTTATTTATTTTTTTTTTGAGGTGGAGTATCGCTATGTCACCCAGGCTGGAG 6728

Qy        458 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 517
              ||||||     |     || |||||||| |||  |||| ||    |||||   |   |
Db       6727 TGCAGTGGCATGATATCAGCTCACTGCAACCTTCACCTGCTAGATTCAAGCAGTTCTCGT 6668

Qy        518 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 577
              ||||||||||  |||||||| || |||  ||     |||||| ||   |||||  |
Db       6667 GACTCAGCCTCCTGAGTAGCTGTGATTACAGGCACATGCCACCATGCTCAGCTAACT-TT 6609

Qy        578 AAAATTTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 637
              |||||||  |||||||| |||||| || | ||||||||||||  ||||  ||||||||
Db       6608 TGTATTTTTTAGTAGAGACAGGGTCTCACCATGTTGGCCAGGCTGGTCTGGAACTCCTGA 6549

Qy        638 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 697
              | |||||  |||| ||  |||||||  |||||||||||||||||||||||||  ||||||||
Db       6548 CCTCAAGTGATCCGCCCTCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACT 6489

Qy        698 ATGTGCGGCTACTTATTTCTTT--------ACATTCCATCTTTCCAATAGAATGTAAGAT 749
                | | ||  || | ||||||          | | |  | || | ||          ||  |
Db       6488 GCGCCCAGCCTCTGAGTTCTTTTAAGTCAGAAACTTGAGCTCTTCAGCTCTGATAAATTT 6429

Qy        750 CCACAGAACAGGGATTACTGCCT--ATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 807
                | |  |||| || |   |   |||||||  | | ||||||  |||  ||||||  |
Db       6428 GGGGGGCAAGGGGACTAATTTTTTCTTTTTCTTTTTCTTTTTTTTAAGATGGAGTCTTGC 6369

Qy        808 TTCATCACCTCAACCTCCGTTCA-------------GCTCACTGCAACCTCTGCCTCCC 853
              |   |  ||          || ||            ||||||||||||||||||||||
Db       6368 TGTGTTGCCCAGGGTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACCTCTGCCTCCC 6309

Qy        854 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 913
              |||||||||: |||||||||||| ||| | |||||||||||| | | || || ||  |||
Db       6308 AGGTTCAAGCAATTCTCCTGCCTCAGCTTACTGAGTAGCTGGGACTGCAGGCCTGTGCCA 6249

Qy        914 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 973
              |||  ||  ||||| |||||||||||||||  |||||  ||||||  |  ||  ||| |||||
Db       6248 CCACTCTCAGCTAA-TTTTTGTATTTTTAGTAGAGACAGGGTTTCAGCACATTGGCCAGG 6190

Qy        974 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1033
              ||||||||  | |||||||||||||||||||||||||  ||||      ||||||||||||  |
Db       6189 CTGGTCTCGATCTCCTGACCTCAAGTGATCTGCCTACCTCGTCTCCCCAAAGTGCTGGGA 6130

Qy       1034 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1093
              |||  ||||  |||  |||     |||||||  |     |   || || ||| ||     |
Db       6129 TTACAGGCAAGAGCCACCACACCTGGCCCTT---GGGGGATGTTATTTCTTTGACAAGTT 6073

Qy       1094 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1153
                | |  |    | ||||   | |  |    |  |  |   || ||       ||
Db       6072 TTGCCTTTCAAATTATATCTGTTGTCTCTTTCAGGAACTCTGTTTAGTTATATTTTGGGT 6013

Qy       1154 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1213
              ||||||  |  ||||  |||| |    |   | |   || |   |||||    | ||
Db       6012 CTTCTAGATTAATCCTTTAATTTTTTAAAATATCTATACGGTTCATCTCTTTGGCAATTA 5953

Qy       1214 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1273
                | | |    || | | ||      ||||  |     | | | ||      | |  |
Db       5952 GTTCTACTTTATACTTTTTCCTTAATTTTTATTTTCCAACTCTTATTTAAATTTCTGTCA 5893

Qy       1274 TCAAACCTTTATT-----------GTTACTGTTTTTTCATTGTTACTATTTTAGTC---- 1318
              |       || |||           ||  | |||||||| |||| || ||||  ||
Db       5892 TATTTTGTTCATTTCTAAGAGTTATTTCATATTTTTTCACTGTTCCTTTTTTTTTCTTTA 5833

Qy       1319 ---------ACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGA 1369
                      | ||| || | |  |||  ||        |||  |  | |     |  | |
Db       5832 GGCTAGTCAAGTGATTATTGTTCCTTTTTTAATAGTGTCATATTGTTTCAGGGATACAAA 5773

Qy       1370 A-------------------------------GGCCAAATTTACAAATCTGATGAAAGC 1397
```

```
                |                                ||   | |||| |  |  |    ||
Db       5772 ATCTCTTACCTTTCTAAGAATTGATTATACTTGGGTATATTTTATATTTAGGGGAGAATA 5713

Qy       1398 TATGAACCCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTT 1457
                  |     | |       | | |                          |     |  |||||
Db       5712 GGGGTTTTTTGTTTGTTTTGTATTGTTGCTGTTGTTTTGTTGTTCAGTTCTCCTGTTTTT 5653

Qy       1458 TAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAA 1517
              |||  ||| |   |  |  |          |  |     ||| | |||  |    ||
Db       5652 TAACTTTTTCTGTTTTGTAATTTTGTTCTCTGTCTGTCATGGTAGTTCTCAAATGTTTTG 5593

Qy       1518 AATAACTCAGTTGGGCACAGTGACTCA-------------------------------- 1544
                    || ||||  ||||||||  |
Db       5592 TGGTCTTCGGTTGTCCACAGTGAAAAATTGTTTTAAAGCACTTGGGGCAGAGCTTATAAA 5533

Qy       1545 ----------AGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAG 1594
                       | ||||||| | ||| |||||||  ||  |  |||  |||  |||||  ||||
Db       5532 CCAACGGATTTGGCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCA--TGAG 5475

Qy       1595 GTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC- 1653
              || ||  ||  |||||||||  |||||   ||  |  ||||||||||  ||||||||||||||||
Db       5474 GTCAGGAGATCGAGACCATCCTGGCTAACACAGTGAAACCCTGTCTCTACTAAAAATACA 5415

Qy       1654 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1713
              ||||||||||  ||| ||  |||  |    ||  |||||||||||||||||||||||||||
Db       5414 AAAAATTAGCCGGGCGTGGTGGCGGGCACCAATAGTCCCAGCTACTCGGGAGGCTGAGGC 5355

Qy       1714 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1773
              | ||||||  ||  |||||||||||||  ||| ||||||||||||  |||||  ||||||||||
Db       5354 AGGAGAATGGCGTGAACCTGGGAGGCGGAGCTTGCAGTGAGCCAAGATCGCACCACTGCA 5295

Qy       1774 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGG 1833
              ||  |||||  ||| |||  |||||||||  | ||  |||||||  ||  ||  ||||||||
Db       5294 CTTGAGCCTAGGCAACAAAGCGAGACTGTCTCAAAAAAAAAATAAAAATAAAAATAAACC 5235

Qy       1834 ATCGGA 1839
                |  |||
Db       5234 AATGGA 5229
```

```
RESULT 5
CD557847/c
LOCUS       CD557847               897 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  AGENCOURT_14423258 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30390116 5', mRNA sequence.
ACCESSION   CD557847
VERSION     CD557847.1  GI:31583915
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM464  row: k  column: 21
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .897
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30390116"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             15.8%;  Score 315.8;  DB 6;  Length 897;
  Best Local Similarity   68.1%;  Pred. No. 4.4e-19;
  Matches  469;  Conservative    1;  Mismatches  213;  Indels    6;  Gaps    2;

Qy        383 TGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTG 442
              |||| |        |  ||   || |||| | ||  | ||| |||||| |||||  |||||
Db        765 TGTATGCATCCCGATATTATTTTATTTTATTTTTATTTTTGAGACATGGTCTCACTCTG 706

Qy        443 TCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGC 502
              |  ||||| |  |||  |||||     | |  || |||||||| |||| ||||||    |
Db        705 TTGCCCAGCCTGGAGTGCAGTGGCAGGATCTCAGCTCACTGCAACCTCCACCTCCCAGGT 646

Qy        503 TCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCAC 562
              |||||  ||   |  |||||||||||||||||||||||||| |||  | ||||||||||||
Db        645 TCAAGCAATTCTCCCACCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCAC 586

Qy        563 GCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTT 622
              |||  |||||||         ||||  ||| |||  |||| || | | | ||||||||||
Db        585 GCCCAGCTAATTTTTTTGTATTTTAAGTACAGATGGGGTTTGCCATGCTGGCCAGGCTC 526

Qy        623 GTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTA 682
              ||| ||| ||||| | |||||  |||  || ||||||||| ||||||||||||||||||||
Db        525 CTCTCAAATTCCTGACCTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGCTGGGATTA 466

Qy        683 CAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAAT 742
              |||| |||||||||| ||  | ||| | |  |  | |  | | | |       | | || |
Db        465 CAGGCGTGAGCCACCGTGCCCAGCTGCATCCTGATATTTAATGCTTGCCAGAATTTGATT 406

Qy        743 GTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGT 802
              | |  || |  |          |  |   | |||| ||     |       | |  |     |
Db        405 GCACCATGCTTATTGGCTTCTATCTTTTTTTTTTTTTTGAGATGGAGACTCACTCTCTCT 346

Qy        803 CTCACTTCATCACCTCAACCTCCGT----TCAGCTCACTGCAACCTCTGCCTCCCGGGTT 858
              ||| |    |   | |      ||     |||||||||||||||||  | ||| |||||
Db        345 CTCCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCCACTTCCTGGGTT 286

Qy        859 CAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATG 918
              ||||:|||||||||| || ||||||| |||||||||| | ||||   |||  ||||||
Db        285 CAAGCGATTCTCCTGTCTCAGCCTCCAGAGTAGCTGGGACTACAGATGTGTGCCACCACA 226

Qy        919 CTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGT 978
              | ||||||||  ||| ||||||||||| |||||||||||||| |||||||||  |||| ||  |
Db        225 CCTGGCTAA--TTTCGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGACTAAT 168

Qy        979 CTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATA 1038
              ||| |||||||||||||| ||||||  ||||||||   | ||||||||||| |||  ||| |
Db        167 CTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGTTGGGGTTACA 108

Qy       1039 GGCGTGAGTCACTGTGCCTGGCCGATTAC 1067
              |||||||| |||    || | | | || |
Db        107 GGCGTGAGCCACCACCCCAGCCGGCTTCC 79


RESULT 6
CR858326
LOCUS       CR858326               3095 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA; cDNA DKFZp469P012 (from clone DKFZp469P012).
ACCESSION   CR858326
VERSION     CR858326.1  GI:55727613
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 3095)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
            Osanger,A., Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp469P012) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp469P012
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3095
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp469P012"
                     /tissue_type="kidney"
                     /clone_lib="469 (synonym: pkid1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="hypothetical protein (Homo sapiens)"
     gene            1. .3095
                     /gene="DKFZp469P012"
     CDS             120. .533
                     /gene="DKFZp469P012"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH90562.1"
                     /db_xref="GI:55727614"
                     /translation="MLSGRLVPGLVSMAGRVCLCQGSAGSGAVGPVEAAIRTKLEQAL
                     SPEVLELRNESGGHAVPPGSETHFRVAVVSSRFEGLSPLQRHRLIHAALAEELAGPVH
                     ALAIQARTPAQWGENSQLDTSPPCLGGNKKTLGTP"
ORIGIN

  Query Match             15.6%;  Score 311;  DB 4;  Length 3095;
  Best Local Similarity   56.6%;  Pred. No. 5.2e-19;
  Matches  883;  Conservative    1;  Mismatches  511;  Indels  166;  Gaps   10;

Qy         384 GTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGT 443
               || ||     ||   | | ||||||| |||  | |  ||| |||| | ||||  || |||
Db        1570 GTGAGCCACTACCTGTAGCGTTTTTGTTTTGTTTGTTTTTGAGATGGAGTCTCACTTTGT 1629

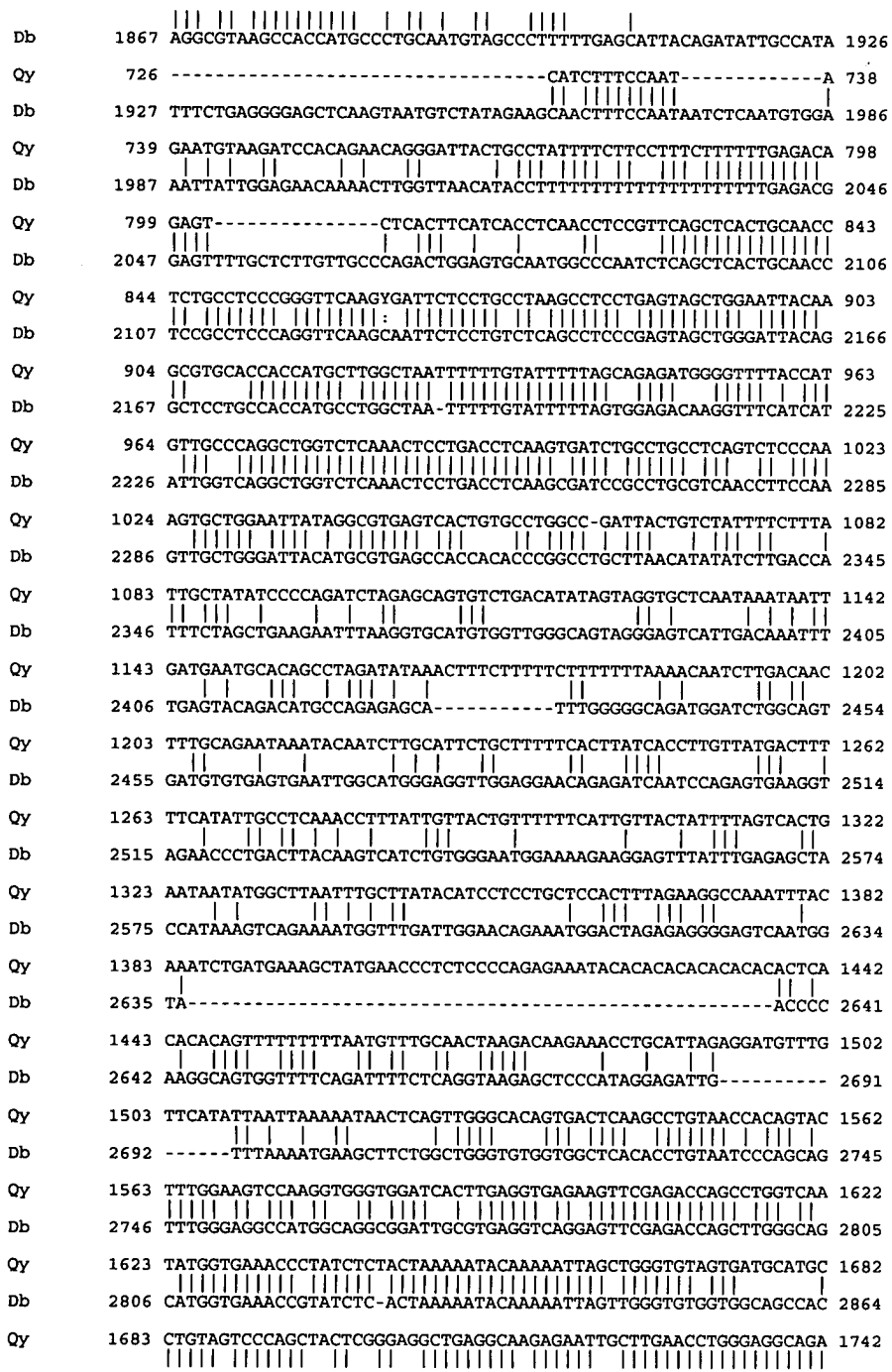
Qy         444 CACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCT 503
                ||||  |  |||  ||||||||| | |  || ||||||||||||  |||||   | |
Db        1630 TGCCCAACCTGGAGTGCAGTGGTGCGATCTCAGCTCACTGCAGCCTCCGCCTCCCAGGTT 1689

Qy         504 CAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACG 563
               ||||  ||   | |||||||||||||||||||||||||||    || ||  |||||||
Db        1690 CAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTAGAGGCTTGTGCCACCACA 1749

Qy         564 CCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTG 623
               ||  |||||||          | |||| | ||||| |||| | || | |||| |||||||  |
Db        1750 CCCAGCTAATT---TTTTGTATTTTATGGAGACAGGGTTTCACCATGTTGCCCAGGCTGG 1806

Qy         624 TCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTAC 683
               ||||| |||||||   ||| ||| ||| || ||||  || |||||| ||||| |||||||
Db        1807 TCTTAGACTCCTGAGCTCAGGCACTCCGCCCACCTCAGCCTCCCAAGGTGCTAGGATTAC 1866

Qy         684 AGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC---------------------- 725
```



RESULT 7
CR749233/c
LOCUS CR749233 5797 bp mRNA linear HTC 19-AUG-2004
DEFINITION Homo sapiens mRNA; cDNA DKFZp686C0331 (from clone DKFZp686C0331).
ACCESSION CR749233
VERSION CR749233.1 GI:51476197
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5797)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSRTM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp686C0331) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686C0331 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .5797
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686C0331"
/tissue_type="testis"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="hypothetical protein"
gene 1. .5797
/gene="DKFZp686C0331"
CDS 103. .1725
/gene="DKFZp686C0331"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH18089.1"
/db_xref="GI:51476198"
/translation="MGPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLG
ITVSKPDLITCLEQGRKPLTMKRNEMIAKPSVMCSHFAQDLWPEQSMKDSFQKVVLRR
YEKCEHDNLQLKKGCISVDECKVHKEGYNELNQCLTTTPRKICQCDKYVKVLHQFPNS
NGQKRGHTGKKPFKYIECGKAFKQFSTLTTHKKIHTGGKPYKCEECGKAFNHSCSLTR
HKKIHTGEKPYKCEECGKAFKHSSTLTTHKRNHTGEKPYKCDKCGKAFMSSSTLSKHE
IIHTEKKPYKCEECGKAFNRSSTLTTHKIIHTGEKPYKCEECDKAFKYSYTLTTHKRI
HTEDKPYKCEECGKAFKYSSTLTTHKRIHTGEKPYKCEECGKAFKRSSDLTTHKIIHT
GEKPYKCEECGKAFKYSSNLTTHKKIHTGERPYKCEECGKAFNQSSILTTHKRIHTGE
KFYKCEECGKAFKRSSHLTTHKIIHTGEKFYKCEECDKAFKHFSNLTTHKKIHTGKKP

YRCEECGKAFKHSSKLTTHKRFHTGEKPYRCE"

ORIGIN

Query Match 15.4%; Score 307.2; DB 4; Length 5797;
Best Local Similarity 56.9%; Pred. No. 7.4e-19;
Matches 902; Conservative 1; Mismatches 569; Indels 114; Gaps 14;

```
Qy       405 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 464
Db      4976 TTTTTTCTTTTTTTTTTTTGTGACAGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTG 4917

Qy       465 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 524
Db      4916 GCACGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTACACCATTCTCCCG-CCTCAG 4858

Qy       525 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 584
Db      4857 CCTCTGGAGTAGCTGGGACTACAGGCGCCCACCACCACGCCCAGCTAATTTA------TT 4804

Qy       585 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 644
Db      4803 TTTAGTAGAGACAGGGTTTCAACGTGTTAGCCAGGATGGTCTTGATCTCCTGACCTCATG 4744

Qy       645 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 704
Db      4743 --ATCCACCTGCCTGGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTCCCTG 4686

Qy       705 GCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGAT 764
Db      4685 GCTTTTTTCTTTCTT----------------------TTTGAGATGGAGTGTCCCTCTGT 4648

Qy       765 TACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTC 824
Db      4647 CACCCAGGCTGGAGTGCAGTGGTGCTAT------------------------------ 4620

Qy       825 CGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 884
Db      4619 --CTCAGCACACTGAAACCTCCACCTCCCAGGTTCAAGCCATTCTCCTGCCTCGGCTTCC 4562

Qy       885 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 944
Db      4561 CGAGTAGCTAGGAATAC--GGGTGCACCACCATGCCCAACTAATTTTTGTTGTTTT---- 4508

Qy       945 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1004
Db      4507 -AGTAGAGAGTTTCACAATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCA--TGATCT 4451

Qy      1005 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGAT 1064
Db      4450 GCCCGACGTCACCTACCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGTCCAGCCCA- 4392

Qy      1065 TACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTA 1124
Db      4391 -----TATTCTATTTATATTTCTGTATAATTTTTATTATGACCATAAAAATCACCCTGTA 4337

Qy      1125 GGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTT 1184
Db      4336 GTCAATTACAATTTAATTGGACATTTAAAAG--TAACTAAAAGTGTATAATTACACTGTT 4279

Qy      1185 TAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTA 1244
Db      4278 TGGAATAGAAAGGATAAATGCTAGAGGTGATAGATACCTTATTTACCCTAATGTAATTAC 4219

Qy      1245 TCACCTTGTTATGACTTTTTCATATTG-----CCTCAAACCTTTATTGTTACTGTTTTTT 1299
Db      4218 TACATATTGTATGCCTGAATCAAAATATTCTCTATAAGACATATATACACATACTATATA 4159

Qy      1300 CATTGTTACTATTTTAGTCACTGAATAATATGGCTT---AATTTGCTTATACATCCTCCT 1356
Db      4158 CTCACAAATAGTAATAATAAATTAAGAAAAAAGATTAAAAATTTAACTAACGGGAACAAT 4099

Qy      1357 GCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGA 1416
Db      4098 ATTCTTCATTTGCAGTTTACTGGCAAAGTGATTACTAGTGATGACATTCCACTACATACC 4039

Qy      1417 GAAATACACACACACACACACTCACACACAGTTTTTTTTTAATG-------------- 1462
Db      4038 AAATAGTATACTTCTTCCATCTTTTGCTTACACCATTTGAGTAAGGCCGGATAGGTTAAA 3979

Qy      1463 ------TTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAA 1516
Db      3978 GTTAGTGGCGTAATAATGCTTCATTAAATGCACAATAGTTTTAAGATGGCAAAAAATAAA 3919

Qy      1517 AAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAG 1576
Db      3918 ATAAAATTAGGCTGGACATATTGGCTCACACCTATAATCCCAACACTTTGAGAGCCCGAG 3859

Qy      1577 GTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCT 1636
Db      3858 GTGGGTGGATCGCTTGAGCTCAGGAGTTTGAGACTAGCCTGACCAATATGGGGAAACCCC 3799

Qy      1637 ATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCT 1696
Db      3798 ATGTCTATTAAAAATATAAAAATTAGCCAGGCGTGGTGTCACGCACCTGTAGTCCCCACT 3739

Qy      1697 ACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAG----TG 1752
Db      3738 ACTCGGGAGGCTGAGGCAGGAGAATTTCTTGAACTTGGGAGGTGGAGGTTGCAGTCATTG 3679

Qy      1753 AGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAA 1812
Db      3678 TGCTGAGACTGCACTGCTGCCCTCCAGCCTGGGTGACAGAGCAAGACTCCATCCCAAAAG 3619

Qy      1813 AATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAA 1872
Db      3618 AAAAGTTAAGTAAATAAAATATAAGTTAAGTTCACACATAATCTAACAATTTTTAAATAT 3559

Qy      1873 GCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTG 1932
Db      3558 ATTGCATTTTATTACGTAACAGTAAAATTAGTAAAATAATTGATATACTTTAATTTATAA 3499

Qy      1933 TTTATTTTGGTTATAAAAGTAAACAA 1958
Db      3498 TTTCTTTCTCTCAGTATAATGTAGAA 3473
```

RESULT 8
CR860521/c

LOCUS	CR860521 1605 bp mRNA linear HTC 12-NOV-2004
DEFINITION	Pongo pygmaeus mRNA; cDNA DKFZp468P2419 (from clone DKFZp468P2419).
ACCESSION	CR860521
VERSION	CR860521.1 GI:55731876
KEYWORDS	HTC.
SOURCE	Pongo pygmaeus (orangutan)
ORGANISM	Pongo pygmaeus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE	1 (bases 1 to 1605)
AUTHORS	Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSRTM	The German cDNA Consortium
TITLE	Direct Submission
JOURNAL	Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp468P2419) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468P2419 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.

```
FEATURES             Location/Qualifiers
     source          1. .1605
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp468P2419"
                     /tissue_type="heart"
                     /clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="unclassified"
ORIGIN

  Query Match             15.1%;  Score 301.2;  DB 4;  Length 1605;
  Best Local Similarity   67.7%;  Pred. No. 5.6e-18;
  Matches  474;  Conservative    1;  Mismatches  204;  Indels   21;  Gaps    3;

Qy        398 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 457
Db       1602 TTTTTTTTTTTTTGTTTGTTTGTTTTGAGACAGAATCTTGCCTTGTCTCCCGGGCTGGAG 1543

Qy        458 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 517
Db       1542 TGTAGTGGTGCAATCTCAGCTCACTGCAACATCCATCTCCCGGATTCCACTTATTCTCCT 1483

Qy        518 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 577
Db       1482 GCCTCAGCCTCTCAGGTAGCTGGGATTATGGGTGCCTGCCATAATGCCCAACTCAATTTT 1423

Qy        578 AAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 637
Db       1422 GTA--CTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGG 1365

Qy        638 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 697
Db       1364 CCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGACCACTG 1305

Qy        698 ATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAA 757
Db       1304 CACCTGGCCTGAATTTCTCCATTCTTCCCA-----CACACCCGCCTCAGGTTCTCCTTCC 1250

Qy        758 CAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCT 817
Db       1249 TGATCATTGACGCTTTTTTTTTTTTTTTTTTTTTGAGACAGCATCTCACTGTCTCACCC 1190

Qy        818 CAACCTCCGTTCA--------------GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGY 863
Db       1189 AGACTGGAGTGCAGTAGTGCAATCTTGGCTCACTGAAACCTCTTCCTCCCAGCCTCAAGC 1130

Qy        864 GATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGG 923
Db       1129 GATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACATGTGTGCACCACTAACACCCA 1070

Qy        924 CTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAA 983
Db       1069 GATAATTTTTGTACTTTTAGTAGAGATGGGGTTTGACCATATTGGCCAGACTGGTCTTGA 1010

Qy        984 ACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGT 1043
Db       1009 ATTCCTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTGTTGGGGTTATAGGCGT 950

Qy       1044 GAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTAT 1083
Db        949 GAGCCACTGCACCCGGCCCCCTTCCTTCATCTTAGTCAAT 910


RESULT 9
CR599842
LOCUS       CR599842               1863 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DA002YO06 of Neuroblastoma of Homo
            sapiens (human).
ACCESSION   CR599842
VERSION     CR599842.1  GI:50480649
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1863)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1863)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1863
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA002YO06"
                     /tissue_type="Neuroblastoma"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             15.1%;  Score 301.2;  DB 4;  Length 1863;
  Best Local Similarity   56.9%;  Pred. No. 5.1e-18;
  Matches  811;  Conservative    1;  Mismatches  509;  Indels  105;  Gaps   10;

Qy        424 GAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTG 483
Db        496 GGGTCAGAGTCTCACTCTGTTTCCCAGGCTGGAGTTCTGTGGCACCGTGTTGGCTCACTG 555

Qy        484 CAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGAC 543
Db        556 CAACCTCCGTCTCCTGGGTTCAAGCAATTCTCTTGCCTCAGCCTCCAGAGTAGCTGGGAT 615

Qy        544 TACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCT 603
Db        616 TACAGGTGCCCACCACCACACCCGGCTAATTTTTTTATATTTTTAGTAGACACGGGGTTT 675

Qy        604 TACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCA 663
Db        676 CACCATGTTGGCCAGGCCGGTCTTGAACTTGTGACGTCAGGTGATGCACCCGCCTTGGCC 735

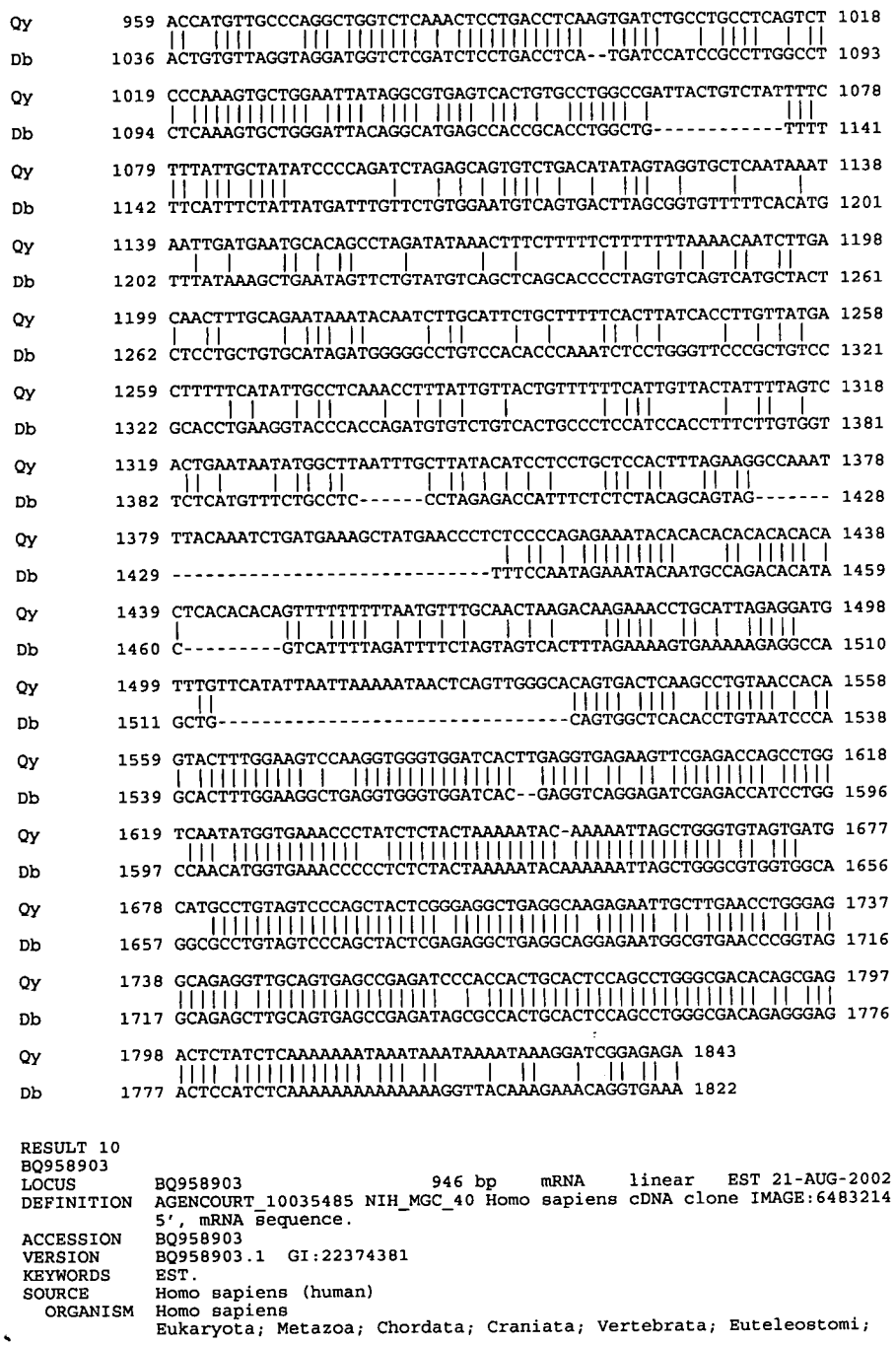
Qy        664 TCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACAT 723
Db        736 TCCCAAAGTGCTGGGATTACAGGTCCAGATTTTGAACAAATCTCGCTAGAGATTGAACAG 795

Qy        724 TCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCT 783
Db        796 AGCTGCGCCTGGATCCCGCCACAGTGAGGAGACCTGAAGACCAGAGAAAACACAGCAAAC 855

Qy        784 TTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAA-C 842
Db        856 GGAGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAGTGTGGCGATCTCGGCTCCCTGCAAGA 915

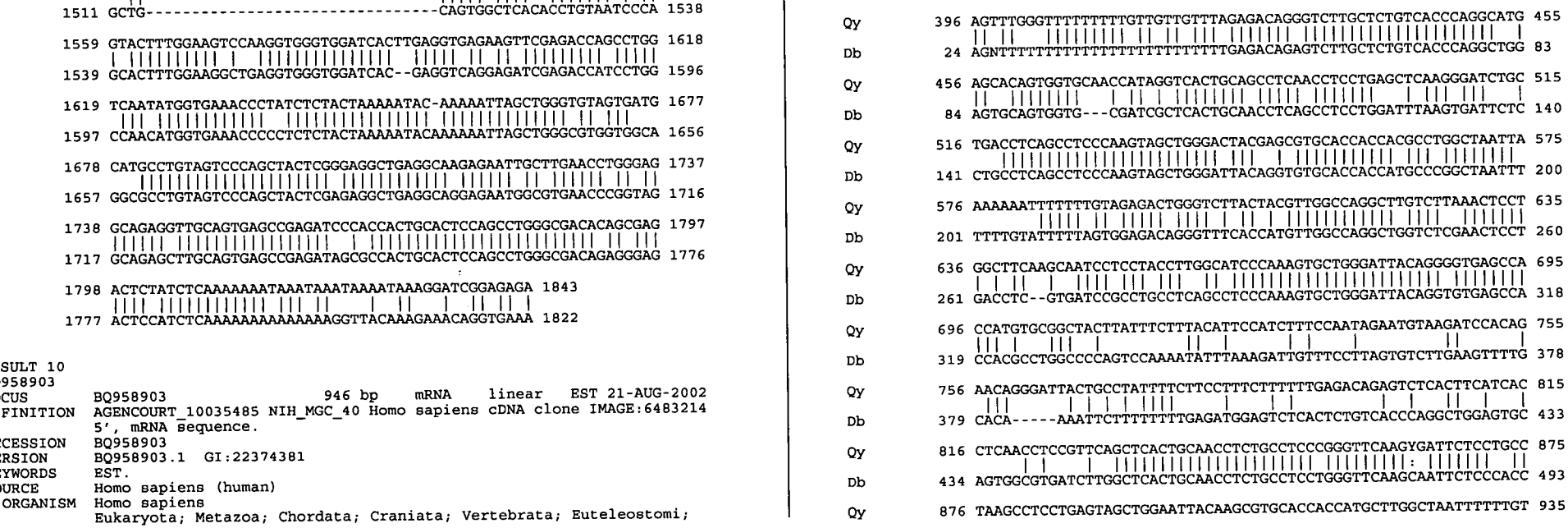
Qy        843 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 902
Db        916 ATCTGCCTCCCAGATTCACGCCATTTTCCCACCTCAGCCCCTGGAGTAGCTGGAAGTACA 975

Qy        903 AGCGTGCACCACCATGCTTGGCTAAT----TTTTTGTATTTTTAGCAGAGATGGGGTTTT 958
Db        976 GGCTCCCGCCACCACGCCTGGCTAATTTTGTTTTTGTATTTTTAGTAGAGACGGGGTTTC 1035
```



```
RESULT 10
BQ958903
LOCUS       BQ958903                946 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10035485 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483214
            5', mRNA sequence.
ACCESSION   BQ958903
VERSION     BQ958903.1  GI:22374381
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 946)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2667  row: g  column: 23
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1. .946
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6483214"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_40"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           15.0%;  Score 300.6;  DB 5;  Length 946;
  Best Local Similarity   69.1%;  Pred. No. 8.9e-18;
  Matches  468;  Conservative    1;  Mismatches  197;  Indels   11;  Gaps    4;
```



```
Db        494 TCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTTGC 552

Qy        936 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 995

Db        553 ATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGGTGGTCTTGAACTCCTGACCTC 612

Qy        996 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1055

Db        613 AGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGC 672

Qy       1056 CTGGCCGATTACTGTCT 1072

Db        673 TCAGCCGCANAATTCCT 689


RESULT 11
BQ706343/c
LOCUS       BQ706343                 918 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8484983 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301091
            5', mRNA sequence.
ACCESSION   BQ706343
VERSION     BQ706343.1  GI:21845242
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2516  row: g  column: 12
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6301091"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             14.9%;  Score 298;  DB 5;  Length 918;
  Best Local Similarity   68.6%;  Pred. No. 1.5e-17;
  Matches  456;  Conservative    1;  Mismatches  197;  Indels   11;  Gaps    3;

Qy        398 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 457

Db        666 TTTTGGATTATGATTATTATTCTTTGAAGCCAGGGTCTCGCTCTGTCACNCAGGCCGGAG 607

Qy        458 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 517

Db        606 TGCAGTGGTGCAATCATGGCTCACTCCAGCCTTTATCTCCTGGGCCCAAGTGATCCTCCA 547

Qy        518 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTA------ 571

Db        546 ACCTCAGCCTCCCAAGTAGCTGGGATTATAAGTGCGTGCTACGATGTCTGGCTAATTTTT 487

Qy        572 -ATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAA 630

Db        486 TATTTATAAATATTTTTTGTAGAGACAGGGTCTCGCTTTGTTGCCCAGGCTGGTCTCAAA 427

Qy        631 CTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTG 690

Db        426 TTCCTGGACTCAGGCGATCCTCCTGCTTCAGCCTCCCAAAGTGCTGAGATTACAGGCGTG 367

Qy        691 AGCC---ACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 747

Db        366 AACCTTTCAACTTTTAATGCTTTCAGTATTTTCCCATCCCTTGATTGAACATTTGGGTTT 307

Qy        748 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 807

Db        306 TAAATTATTTTTGTTGTTGTTGTTTCTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGC 247

Qy        808 TTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATT 867

Db        246 TAGAATGCAGGGGTGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGCGTTCAAGCGATT 187

Qy        868 CTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA 927

Db        186 CTTGTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCTGCCATGATGCCCAACTAA 127

Qy        928 TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTC 987

Db        126 TTTTTTGTATTTTTAGTAGAGACAGTGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTC 67

Qy        988 CTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCA-AAGTGCTGGAATTATAGGCGTGAG 1046

Db         66 CTGACCTCAAGTGATCCGCGCACCTTGGCCTCCCANAATTGCTGGGATTACAGGTGTGAG 7

Qy       1047 TCACT 1051

Db          6 CCACT 2


RESULT 12
CX780936
LOCUS       CX780936                 736 bp    mRNA    linear   EST 02-MAR-2005
DEFINITION  HESC3_2_B04.b1_A036 NIH_MGC_260 Homo sapiens cDNA clone
            IMAGE:30928759 3', mRNA sequence.
ACCESSION   CX780936
VERSION     CX780936.1  GI:58297726
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: HESC3_2_B04.g1_A036
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: BresaGen, Inc.
             cDNA Library Preparation: Express Genomics, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
            University of Georgia
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
          Plate: NDAM1172  row: c  column: 08
          Seq primer: M13-21 (TGTAAAACGACGGCCAGT)
          High quality sequence stop: 711
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30928759"
                     /sex="male"
                     /tissue_type="human embryonic stem cells"
                     /cell_type="human embryonic stem cells"
                     /cell_line="BG01"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="NIH_MGC_260"
                     /note="Vector: pExpress-1; Site_1: NotI; Site_2: EcoRV;
                     RNA obtained from human embryonic stem cells isolated from
                     the inner cell mass of blastocyst stage embryos. Cell line
                     id and NIH Registry designation is BG01. Positive for
                     SSEA3, SSEA4, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and
                     Nanog expression; negative for SSEA1 expression. Passage
                     number 21. cDNA primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. This primary library
                     is non-normalized (normalized primary library is
                     NIH_MGC_261). It was constructed by Express Genomics
                     (Frederick, MD). Sequence ends have been trimmed to
                     exclude vector and regions below Phred quality 16. Note:
                     this is a Mammalian Gene Collection library."
ORIGIN

  Query Match             14.7%;  Score 293.6;  DB 8;  Length 736;
  Best Local Similarity   66.9%;  Pred. No. 4.3e-17;
  Matches  464;  Conservative    1;  Mismatches  220;  Indels    9;  Gaps    3;

Qy        399 TTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 458
              ||   ||||||||| || || ||| ||||  | ||||  |||| |  |||||||   ||
Db          1 TTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTCTTGCCCAGGCTGAAGT 60

Qy        459 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 518
              |||||| |  | |   | |||||||| ||||  |||||||  | |||||   |
Db         61 GCAGTGGCGGGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCACTTCTTCTG 120

Qy        519 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 578
              | |||||| ||  |||||||||||| ||| ||   |||||||| |||||| || ||
Db        121 CTTCAGCCCCCTGAGTAGCTGGGATTACAGGCACCCACCACCAGGCCTGGATACTT--TT 178

Qy        579 AAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGC 638
                | ||||||    | | ||||  || |||  |||||||||||| |||| |||||||||  |
Db        179 TTGTATTTTTAGTAAAAATGGGGTTTGCTATATTGGCCAGGCTGGTCTCAAACTCCTGAC 238

Qy        639 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA 698
              ||| |  |||| ||| |||  || | ||||||||||||||||||||  ||| ||||||
Db        239 CTCAGGTGATCCACCTGCCTAAGCCTGCCAAAGTGCTGGGATTACAGGCATGAACCACCA 298

Qy        699 TGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAAC 758
              ||  | ||  |||  |||||  | |                ||  | |  |  |
Db        299 TGCCCAGCCTCTTTATTCTTTTTAATATTATCACAAGTTATTTATTTATTTATTTATTTA 358

Qy        759 AGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTC 818
                |||  |   |     |     |||    ||   | |  ||    | | ||    | |
Db        359 TTTATTTATTTTTTGAGATGGAGTTTTGCCCTTGTTGCCCAAGCTGGAGTGCAATGGCAC 418

Qy        819 AACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAA 878
              | |    |||||||||||||||||  ||||||||||||||:||||||||||||| |
Db        419 GATC-----TCAGCTCACTGCAACCTCCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCA 473

Qy        879 GCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATT 938
              |||||||||||||||| |||||| ||     ||| ||||| |||||||  |||||  |||
Db        474 GCCTCCTGAGTAGCTGGGATTACAGGCACATGCCATCATGCCTGGCTAA--TTTTGCATT 531

Qy        939 TTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAG 998
```

```
              ||||| |||||  | ||||  ||||||||  ||||||||||  |||||||||||||| |
Db        532 TTTAGTAGAGACAGAGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGG 591

Qy        999 TGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTG 1058
              |||||||||  |||| | ||||||||||||||| |||||||||  |||| ||| |||||
Db        592 TGATCTGCCCTCCTCGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCGTGCCCA 651

Qy       1059 GCCGATTACTGTCTATTTTCTTTATTGCTATATC 1092
              |||  |  |      | ||  |  || || | ||
Db        652 GCCATTATCACAAGTTATTTATACTTCCTTTCTC 685


RESULT 13
CR613629
LOCUS       CR613629               3092 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI039YN19 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR613629
VERSION     CR613629.1  GI:50494436
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3092)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 3092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .3092
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI039YN19"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             14.5%;  Score 290.6;  DB 4;  Length 3092;
  Best Local Similarity   69.7%;  Pred. No. 3.1e-17;
  Matches  466;  Conservative    1;  Mismatches  190;  Indels   12;  Gaps    5;

Qy        394 ACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCA 453
              | | |||   |||||||| || || ||| ||||||| ||||||||||||||||||||||
Db       1750 AAATTTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCT 1809

Qy        454 TGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCT 513
              |||  ||||||||   | || | |||||||| ||||| |||||||   | ||| |||
Db       1810 GGAGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAGCCTCCTGGATTTAAGTGATTC 1866

Qy        514 GCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAAT 573
              |   |||||||||||||||||||||||| |||  | |||||||||| ||| |||||||
Db       1867 TCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCATGCCCGGCTAAT 1926

Qy        574 TAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 633
              |    | ||||| || ||||| |||| | || | |||||||||||| ||||  |||||
Db       1927 TTTTTGTA-TTTTTAGTGGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTC 1985
```

```
Qy        634 CTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC 693
              ||| | ||  |  |||| ||| |||  || ||||||||||||||||||||||| ||||||
Db       1986 CTGACCTC--GTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGC 2043

Qy        694 CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC 753
              ||||| |  |||  |          ||  |      | |    |         ||
Db       2044 CACCACGCCTGGCCCCAGTCCAAAATATTTAAAGATTGTTTCCTTAGTGTCTTGAAGTTT 2103

Qy        754 AGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATC 813
              | |||      | | | ||||      |   | |          |  |  ||  |
Db       2104 TGCACA-----AAATTCTTTTTTTTTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGT 2158

Qy        814 ACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTG 873
              |     | |    |  |||||||||||||||||||||||| |||||||||: |||||||
Db       2159 GCAGTGGCGTGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCAATTCTCCCA 2218

Qy        874 CCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT 933
              ||| |||||||  |||||||||| ||||||  ||||  |||||||  | ||| ||| |||||
Db       2219 CCTCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTT 2277

Qy        934 GTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACC 993
              | ||||||||  |||| || |||| |||||||| ||||| ||||||  ||||||||||
Db       2278 GCATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGTTGGTCTTGAACTCCTGACC 2337

Qy        994 TCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGT 1053
              ||| ||||||  |||||||| | |||||||||||| |||| |||| |||| ||| ||
Db       2338 TCAGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGT 2397

Qy       1054 GCCTGGCCG 1062
              ||  ||||
Db       2398 GCTCAGCCG 2406
```

RESULT 14
BQ681302/c
LOCUS BQ681302 855 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8034339 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213772 5', mRNA sequence.
ACCESSION BQ681302
VERSION BQ681302.1 GI:21793981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 855)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2380 row: a column: 05
High quality sequence stop: 649.
FEATURES Location/Qualifiers
source 1. .855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6213772"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 14.4%; Score 288.2; DB 5; Length 855;
Best Local Similarity 69.2%; Pred. No. 1.1e-16;
Matches 471; Conservative 1; Mismatches 161; Indels 48; Gaps 4;

```
Qy        406 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 465
              ||||||| || || |||  |||||||||||||| ||||||||||||  |  |||  ||||||
Db        855 TTTTTTTTTTTTTTTTTTTAGACAGGGTCTTGATCTGTCACCCAGACTGGAGTGCAGTGG 796

Qy        466 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 525
              ||  | ||||| |||||| | | |||| ||| |  |||||  |  ||       || |||||
Db        795 TGAGATCATAGCTCACTGTAACNTCAAACTCTTCGGCTCAGGCAATACTTCCACTTCAGC 736

Qy        526 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTT 585
              |||||| |||||||||||||| || |    ||||||  ||  |||||  |      || |||
Db        735 TTCCCAAATAGCTGGGACTACAAGTGCATGCCACCATACCCAGCTAAAT-GTTTAAGTTT 677

Qy        586 TTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGC 645
              ||||||||||  ||||||| | | |||   ||||||  ||||||||||||||||||||||
Db        676 TTTGTAGAGATGGGGTCTTGCCATGTTTCNCAGGCTGCTCTTAAACTCCTGGCTTCAAGT 617

Qy        646 AATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGG 705
              |||||||  ||||  || |||||||||||||||||  |||||  |||||  |||  |||   ||
Db        616 GATCCTCCCACCTCAGCCTCCCAAAGTGCTGGGATGACAGGTGTGAGACACAATGCCTGG 557

Qy        706 CTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATT 765
              || ||   |   |  |  |   | ||||  | | |  ||
Db        556 CTTCTCTCTCTCTCTCTTTTTTTTTCTGAAACAGAGT---------------------- 518

Qy        766 ACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCC 825
                              ||| || ||   | |   |   || | ||    ||| |
Db        517 -----------------TTTGTTCTTGTTGCCCAGGCTGCAGTGCAATGGCTCGA----- 480

Qy        826 GTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCT 885
              |||||||||||||||||||| ||||||||||||||:|||||||||||||  |||||||
Db        479 TTTCAGCTCACTGCAACCTCTACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCC 420

Qy        886 GAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA----TTTTTTGTATTTTT 941
              |||||||| |||||| |||    |||||| || ||||| |    ||||||||||||||
Db        419 AAGTAGCTGGGATTACAGGCGCCTGCCACCACGCCTGGCTGACTTTTTTTTTGTATTTTT 360

Qy        942 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1001
              || |||||||||||||| |||||| || ||||||||  ||  |||||||| |||| ||||
Db        359 AGTAGAGATGGGGTTTCACCATGGTGGCCAGGCTGACCTTGAACTCCTGATCTCAGGTGA 300

Qy       1002 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1061
              ||  ||  |||  | |||||||||||||||| |||| ||| ||||| |||||||| ||||
Db        299 TCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCCGGCC 240

Qy       1062 GATTACTGTCTATTTTCTTTA 1082
                | ||  |  | ||  | |
Db        239 AGCTTCTTGCCTTCTTACTGA 219
```

RESULT 15
BQ710315/c
LOCUS BQ710315 769 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8351732 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282212 5', mRNA sequence.
ACCESSION BQ710315
VERSION BQ710315.1 GI:21849214
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 769)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2476  row: d  column: 21
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6282212"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             14.4%;  Score 287.6;  DB 5;  Length 769;
  Best Local Similarity   67.0%;  Pred. No. 1.4e-16;
  Matches  453;  Conservative    1;  Mismatches  215;  Indels    7;  Gaps    3;

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
              ||||||||| || || | | ||||||| |||| |||||| |||||||  ||| ||||
Db        768 TTTTTTTTTTTTTTTTTATTGAGACAGAGTCTCACTCTGTCCCCCAGGCTGGAGTGCAGT 709

Qy        464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
              || |  | |   | |||||||| ||||  ||||||| | ||||| |||   |   |||||
Db        708 GGGGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCA 649

Qy        524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
              ||||||| ||||||||||||||   |||  | ||||   ||| |   | ||  | |   |
Db        648 GCCTCCCGAGTAGCTGGGACTATAGGCGCCCGCCACTGTGCCGGCTAATTTTTATA---T 592

Qy        584 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 643
              |||| |||||||| |||| |  | | ||||||||||||| ||||  || ||||  | |||
Db        591 TTTTAGTAGAGACGGGGTTTCCCCATGTTGGCCAGGCTGGTCTCGAATTCCTAACCTCAG 532

Qy        644 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 703
              |  |||| |||  ||  || ||||||  |||||||||||||||| |||||||||||||  |
Db        531 GTGATCCGCCTGGCTCAGCCTCCCAACATGCTGGGATTACAGGTGTGAGCCACCATGCCC 472

Qy        704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 763
              ||||   |    ||||   |    |||  |   |          | | ||      |
Db        471 GGCTGTGTTACACTTTTTGTGGTGTCTCACAGGTCTCTGAGGCTCTGCTCATTTTTGTTT 412

Qy        764 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 823
              ||  |      |||| |          | |                ||| |   |
Db        411 TTGTTTTTATTTTTTTGAGACAGAGTCTCTTTCTGTCACTAGACTGGAGTGCAGTGGTGC 352

Qy        824 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 883
                || |||||||||||||||||||||| | |||||||:||||||| ||||| ||||||
Db        351 AATCTCGGCTCACTGCAACCTCTGCCTCCAGAGTTCAAGCGATTCTCTTGCCTCAGCCTC 292

Qy        884 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA--TTTTTTGTATTTTT 941
              |  |||||||| |||||| ||     |||||||||    ||||  ||||||||||||||
Db        291 CCAAGTAGCTGGGATTACAGGCACATGCCACCATGCCCAACTAATTTTTTTTGTATTTTT 232

Qy        942 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1001
              || |||||||||||||| |||||||||| ||||| ||||||  | ||| |||||||  ||||
Db        231 AGTAGAGATGGGGTTTGACCATGTTGGCCAGGATGGTCTAGATCTCTTGACCTC--GTGA 174

Qy       1002 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1061
              ||   |  |||  | |||||||||||||||  |||| |||||||| ||||| ||||||||
Db        173 TCCAACCACCTTGGCCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACTGCGCCTGGCC 114

Qy       1062 GATTACTGTCTATTTT 1077
              |  | ||   | || |
Db        113 GGCTCCTCCTTCTTAT 98


Search completed: February  9, 2006, 18:10:19
Job time : 9445.43 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 65.2558 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 3524
Sequence: 1 ctatctcagctgtcccctccc.....ckctacataaatatctatta 2001

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_605_@runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	459	13.0	7067	6	AAE35265	Aae35265 Human P45
	2	449.5	12.8	7052	6	AAE35266	Aae35266 Human P45
c	3	444	12.6	7052	6	AAE35266	Aae35266 Human P45
	4	438.5	12.4	215	4	AAU31513	Aau31513 Novel hum
c	5	432	12.3	7036	6	AAE35264	Aae35264 Human P45
	6	427.5	12.1	217	7	ADC87151	Adc87151 Human GPC
	7	420.5	11.9	7067	6	AAE35265	Aae35265 Human P45
	8	416	11.8	673	6	ADA84009	Ada84009 Human POM
	9	413	11.7	185	4	AAO06014	Aao06014 Human pol
	10	405	11.5	232	5	ADK36978	Adk36978 Novel hum
	11	400	11.4	302	5	ADK36936	Adk36936 Novel hum
	12	400	11.4	361	5	ABG68738	Abg68738 Human pro
	13	400	11.4	449	5	ABJ01057	Abj01057 Ovary cel
	14	397.5	11.3	304	4	AAU30391	Aau30391 Novel hum
	15	397.5	11.3	375	3	AAB01399	Aab01399 Neuron-as
	16	397.5	11.3	375	5	ABB81538	Abb81538 Neural th
	17	397.5	11.3	375	5	AAE29142	Aae29142 AD7c-NTP
	18	397.5	11.3	375	6	ABR63268	Abr63268 AD7c-NTP
	19	397.5	11.3	375	6	ADA84017	Ada84017 Human POM
	20	397.5	11.3	375	6	ABU03024	Abu03024 Human neu
	21	397.5	11.3	375	6	ABB99774	Abb99774 Amino aci
	22	397.5	11.3	375	6	ABJ19445	Abj19445 AD7C-neur
	23	397.5	11.3	375	7	ADB37642	Adb37642 Human neu
	24	397.5	11.3	375	8	ADR14409	Adr14409 Human NF-
	25	397.5	11.3	375	9	AEA30020	Aea30020 Human neu
	26	396.5	11.3	239	8	ADQ65417	Adq65417 Novel hum
c	27	391.5	11.1	673	6	ADA84009	Ada84009 Human POM
	28	390	11.1	382	4	AAU32610	Aau32610 Novel hum
	29	390	11.1	382	4	AAU31818	Aau31818 Novel hum
	30	390	11.1	382	4	AAU32707	Aau32707 Novel hum
	31	388	11.0	7036	6	AAE35264	Aae35264 Human P45
	32	387.5	11.0	396	4	AAU30455	Aau30455 Novel hum
	33	386	11.0	179	8	ADR09211	Adr09211 Human pro
	34	384.5	10.9	158	7	ADB64893	Adb64893 Human pro
	35	384	10.9	324	4	AAU29573	Aau29573 Novel hum
	36	384	10.9	324	4	AAU29579	Aau29579 Novel hum
	37	384	10.9	397	2	AAR95913	Aar95913 Neural th
	38	383.5	10.9	257	4	AAO10294	Aao10294 Human pol
	39	381	10.8	241	8	ADQ66812	Adq66812 Novel hum
	40	381	10.8	527	7	ADD27133	Add27133 Human adi
	41	380.5	10.8	175	4	AAU31857	Aau31857 Novel hum
	42	380.5	10.8	250	4	AAU31823	Aau31823 Novel hum
	43	380.5	10.8	250	7	ADE09441	Ade09441 Novel pro
	44	377	10.7	215	4	ABG07921	Abg07921 Novel hum
	45	376	10.7	189	7	ADB65042	Adb65042 Human pro

ALIGNMENTS

RESULT 1
AAE35265
ID AAE35265 standard; protein; 7067 AA.
XX
AC AAE35265;
XX
DT 28-MAY-2003 (first entry)
XX
DE Human P450RAI-3 protein #2.
XX
KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW enzyme.
XX
OS Homo sapiens.
XX
PN WO200295034-A2.
XX
PD 28-NOV-2002.
XX
PF 23-MAY-2002; 2002WO-CA000758.
XX
PR 23-MAY-2001; 2001US-0292531P.
XX
PA (CYTO-) CYTOCHROMA INC.
XX
PI Wisniewski J, Petkovich PM, Ramshaw H;
XX
DR WPI; 2003-112152/10.
DR N-PSDB; AAD53861.
XX
PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT diagnosing, preventing, or treating disorders with aberrant expression or

PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
XX
PS Example 1; Fig 1; 231pp; English.
XX
CC The invention relates to retinoic acid metabolising cytochrome P450,
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX
SQ Sequence 7067 AA;

Alignment Scores:

Pred. No.:	2.6e-43	Length:	7067
Score:	459.00	Matches:	196
Percent Similarity:	40.9%	Conservative:	59
Best Local Similarity:	31.4%	Mismatches:	186
Query Match:	13.0%	Indels:	183
DB:	6	Gaps:	25

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35265 (1-7067)

```
Qy       1977 AAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGT 1918
Db         77 LysValTrpTyrIleHisThrIleTyrTyrTyrLeuSerIleLysArgAsnGluValIle 96

Qy       1917 TAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGA--- 1861
Db         97 MetHisAlaThrThrThrLeuLysIle-----------LeuCysLeuSerGluArgSer 112

Qy       1860 --------------------ATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTAT 1822
Db        113 GlnLeuGlnLysThrArgValPheLeuPheCysLeuPhe-----------CysPheVal 128

Qy       1821 TTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGG 1762
Db        129 LeuPheCysPheLysArgGlnGlyLeuAlaLeuLeuProIleLeuGlu-Tyr------Se 146

Qy       1761 ATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCC 1702
Db        146 rSerLeuLeuThrAlaAlaSerThrPheTrpThrGlnAlaSerLeuProProGlnProPr 166

Qy       1701 CGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTA 1642
Db        166 oLys---LeuAlaLeuGlnValCysThrThrThrAlaSer---PheLeu-IleValCysA 184

Qy       1641 GAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCA 1582
Db        184 rgAsnArgValSerProCysCysProGlyTrpSerGlnThrProSerLeuLys------- 201

Qy       1581 CCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCACTGTGCCCAACTGAGT 1522
Db        202 ----------------------------------------------------ProTrpProL 205

Qy       1521 TATTTTTAATTAATATGAACAAACATCCTCTAATGCAGGTTTCTTGTCTTAGTTGCAAAC 1462
Db        205 euLysValLeuGlyLeuGlnGlyAlaThrMetValGly---------------------- 217

Qy       1461 ATTAAAAAAAAACTGTGTGTGAGTGTGTGTGTGTGTGTGTATTTCTCTGGGGA------- 1409
Db        218 -----GlnLysThrThrCysTyrLeuIleProIle-----------TrpGlyValIleA 232

Qy       1408 ----------------------GAGGGTTCATAGCTTTCATCAGATTTGTAAATTTGGC 1372
Db        232 laAsnGlyTyrGlyValSerPheArgGlySerLysIleAspAspGly-CysThr------ 249

Qy       1371 CTTCTAAAGTGGAGCAGGAGGATGTATAAGCAAATTAAGCCATATTATTCAGTGACTAAA 1312
Db        250 IleLeuIleGluSerLysProLeuIleCysAlaLeuMetSerLeuTyrGlyThrSerGln 269

Qy       1311 ATAGTAACAATGAAAAAAC------------AGTAACAATAAAGGTTTGAGGCAATATGA 1264
Db        270 SerCysLeuLys-LysAsnPheSerTyrIleThrArgAsnArgGlyLeu----------- 285

Qy       1263 AAAAGTCATAACAAGGTGATAAGTGAAAAAGCAGAATGCAAGATTGTATTTATTCTGCAA 1204
Db        285 ---------------------------------------------------------- 285

Qy       1203 AGTTGTCAAGATTGTTTTAAAAAAAGAAAAAGAAAGTTTATATCTAGGCTGT-------- 1152
Db        286 ----------IleIleLeuLysLeuLysGlnLysThrLysMetValGlyLeuProLysLy 302

Qy       1151 ----------------GCATTCATCAATTATTTATTGAGC------ACCTACTATATGTC 1114
Db        302 sArgAlaGlyIleLysGlyLysPheSerTyrValIleAlaValLysThrGlnAsnLysAs 322

Qy       1113 AGACACTGCTCTAGATCTGGGGATATAGCAATAAAGAAAATAGACAGTAATCGGCCAGGC 1054
Db        322 nLysProIleArgAspSerGluIleIleThrArgThrLysArgGlyValLysGlyGlnTh 342

Qy       1053 ACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGA 994
Db        342 rGlnTrpLeuThrProValIleProAlaPheTrpGluAlaGluValGlyGlySerLeuGl 362

Qy        993 GGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATAC 934
Db        362 uAlaArgSerLeuArgProAlaTrpAlaThrTrp---AsnProValSerThrLysAsnTh 381

Qy        933 AAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGG 874
Db        381 r-LysIleSerArgAlaTrpTrpTyrValProValValProAlaAlaTrpGluAlaGluA 401

Qy        873 CAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGGTTGAGGT 814
Db        401 laGlnGluLeuLeuGluPro---------------------------------------- 407

Qy        813 GATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCCCTGTTCT 754
Db        407 ---------------------------------------------------------- 407

Qy        753 GTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTAGCCGCACATGGTG 694
Db        408 ----------------------------------------GluAlaArgCysGlyG 413

Qy        693 GCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGATTGCTTGAAGCCAG 634
Db        413 lySerArgLeu---SerGlnHisPheGlyArgProArgTrpValAsp--HisGluValAr 431

Qy        633 GAGTTTAAGACAAGCC------TGGCCAACGTAGTAAGACCCAGTCTCTACAAAAAAATT 580
Db        431 gSerLeuArgProAlaProThrTrp-----------AsnProIleSer---AsnLysTy 446

Qy        579 TTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCTGAG 520
Db        446 rLysLysLeuAlaGlyHisGlyGlyVal---CysLeuCysGlnLeuLeuArgArgLeuAr 465

Qy        519 GTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCACCACTG 460
Db        465 gGlnGluAsnCysLeuAsnLeuGlyGlyGlyGlyCysSerGluProArgSerArgHisCy 485

Qy        459 TGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAACAACAAAAAAAAACCCA 400
Db        485 sThrProThrTrpAlaThrLysArgAspSerValSerLysLysLysLysLysLysGl 505

Qy        399 AACTGTCACCACTTACAAGCTGTGTGATCTTGGGCAAATAGCTCAGCCTTCTGAGCCTCA 340
Db        505 uLeuLeuGluSerGlyArgArgArgLeuGlnAlaGluIleAla------------ProLe 521

Qy        339 GTTTTCTTATCTGTTAAGTGAGGATAATTGTACCTATCCCATAGTGTTGTTCTGAGGATT 280
Db        521 uHisSerSerLeuGlyAspArgAlaArgLeuCys------------------------Le 533
```

```
Qy        279 AAAGGAGACACTGCCCAGAACCAAGAAAATTAACACTAAATCA---TGGCTGTTGTTA 225
Db        533 uLysThrThrThrThrGluAsnThrLysAsnLysIleArgSerThrTrpLeuLeuLeu 552

RESULT 2
AAE35266
ID   AAE35266 standard; protein; 7052 AA.
XX
AC   AAE35266;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #3.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7052 AA;

Alignment Scores:
Pred. No.:              3.39e-42      Length:          7052
Score:                  449.50        Matches:         182
Percent Similarity:     36.9%         Conservative:    66
Best Local Similarity:  27.1%         Mismatches:      168
Query Match:            12.8%         Indels:          256
DB:                     6             Gaps:            21

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35266 (1-7052)

Qy        532 AGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT---TTTTTT 588
Db          1 SerSerTrpAspTyrArgCysLeuProProCysProAlaAsnPheLeuTyrIlePheLeu 20

Qy        589 GTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAAT 648
Db         21 ValGluThrGlyPheHisCysIleSerGlnAspGlyLeuAspLeuMet---SerSer--S 39

Qy        649 CCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTA 708
Db         39 erThrLeuLeuGlyLeuProLysCysTrpAspTyrArgArgGluProArgCysProAlaG 59

Qy        709 CTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC---------------- 753
Db         59 ln---------GlnHisTyr-PheProLysLysIleAsnValHisGlnLeuMetIleGly 75

Qy        754 ---AGAACAGGGATTACTGCCTATTTTCTT---------------------------- 780
Db         76 ThrArgCysGlyThrSerIleGlnTyrIleIleIleCysGlnLysGlyMetLysCysMet 95

Qy        781 ---------------CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCC 825
Db         96 LeuGlnArgAspGluProLysTyrTyrAlaValLysGlu-AlaSerTyrLysArgProGl 115

Qy        826 GTTCAGC---------------------------------------------------- 832
Db        115 yArgSerPheCysPheValCysPheValLeuCysCysPheValLeuAsnArgGlyArgVa 135

Qy        833 ------------------------------------TCACTGCAACCTCTGCCTCCCGG 855
Db        135 lLeuLeuCysCysProSerTrpSerThrValHisCysSerLeuGlnProProProSerGl 155

Qy        856 GTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACC 915
Db        155 yLeuLysLeuAlaSerHisLeuSerLeuProSerSer---HisTyrArgCysAlaProPr 174

Qy        916 ATGCTTGGCTAATTTTTTGTATTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCT 975
Db        174 oGlnProAlaAsnPheLeuTyr--Val---GluThrGlySerHisArgValAlaGlnAla 192

Qy        976 GGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATT 1035
Db        193 GlyLeuLysLeuLeuValSerSer---------------------------------- 200

Qy       1036 ATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCC 1095
Db        201 LeuGlyLeuSerLysCysAspTyr---------------------------------- 208

Qy       1096 AGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAG 1155
Db        209 ----------------------------------------------------ArgGlyGlu 211

Qy       1156 CCT-----------AGATATAAACTTTCTTTTTCTTTTTTAAAACAATCTTGACAACT 1203
Db        212 ProProTrpAlaArgArgProHisValIlePheProPheGlyGluLeuLeuMetAlaThr 231

Qy       1204 ------TTGCAGAATAAATACAATCTT---------------GCATTCTGCTTTTTCACT 1242
Db        232 GlyPheLeuSerGlyAlaLeuLysLeuIleGluMetValAlaGlnPheCysGluSerGln 251

Qy       1243 TATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCAT 1302
Db        252 AsnHisPheValProPheLysValAsnCysMet--ValHisLeuLysAlaValLysLysI 271

Qy       1303 TGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCA 1362
Db        271 le---------------------------------PheLysValIleGlnGluIleGluV 280

Qy       1363 CTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATA 1422
Db        280 alLeuTyrSerTyrLysAsnLysAsnArgGlnLysTrpAspSerTyrProArgArgGluL 300

Qy       1423 CACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAA 1482
Db        300 euValSerLysValAsnSerHis---MetLeuAsnAsnLysLeuLysIleArgLysIleS 319

Qy       1483 CCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACT 1542
Db        319 erLeuLeuGlyIleValArgSerLeuGluLeuLysGluGluLeuLysAlaArgArgSerG 339

Qy       1543 CAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAG 1602
Db        339 lySerHisLeuSerGlnHisPheGlyArgProArgTrpValAspHisLeuArgProGlyV 359
```

```
Qy       1603 TTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAG 1662
Db        359 al---AspGlnProGlyGlnHisGlyGluThrProSerLeuGlnLysThrGlnLysLeuA 378

Qy       1663 CTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATT 1722
Db        378 laGlyArgGlyGlyMetCysLeu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnC 397

Qy       1723 GCTTGAACCTG---------------------------------------------------- 1733
Db        397 ysLeuAsnLeuArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyG 417

Qy       1733 ---------------------------------------------------------------- 1733
Db        417 lnAspGlyTrpIleThrArgSerGlyValAspGlnProAspGlnHisGlyGluThrProS 437

Qy       1733 ---------------------------------------------------------------- 1733
Db        437 erLeuThrLysAsnThrLysAsnProGlyMetValAlaCysAlaCysAsnAlaSerTyrS 457

Qy       1733 ---------------------------------------------------------------- 1733
Db        457 erGlyGlyGlyArgArgThrAlaThrTrpGluValGluValAlaValSerArgAspHisA 477

Qy       1733 ---------------------------------------------------------------- 1733
Db        477 laThrAlaLeuGlnProGlyGlnGlnSerGluThrProSerGlnLysLysLysLysLysL 497

Qy       1734 --------------------GGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1773
Db        497 ysLysAsnCysLeuAsnProGlyGlyGluGlyCysSerGluProArgLeuHisHisCysT 517

Qy       1774 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGG 1833
Db        517 hrProValTrpValThrGluArgAspSerValSerLysGln------GlnGlnGlnLysT 535

Qy       1834 ATCGGAGAGAAACAAAACTAATAAGA----------------------------TTCC 1863
Db        535 hrGlnLysThrAsnLysGlyValLysHisGlyTyrPheValArgGlnThrValGlyPheM 555

Qy       1864 TGAAGGTAAGCAGAGATACGTAAATTATATGTAATA---------------------- 1899
Db        555 etIleCysSer-IleLeuLeuThrMetTyrLeuIleLysLeuIleLeuIleAsnProHis 574

Qy       1900 AAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAG 1959
Db        575 LysPheThrCysMetTyrLeuTyrAlaIleLeuAlaMetValValThrLysLysLysLys 594

Qy       1960 CCAAAAGTAATGCAACTT 1977
Db        595 IleLysIleLysGlnIle 600
```

RESULT 3
AAE35266
ID AAE35266 standard; protein; 7052 AA.
XX
AC AAE35266;
XX
DT 28-MAY-2003 (first entry)
XX
DE Human P450RAI-3 protein #3.
XX
KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW enzyme.
XX
OS Homo sapiens.
XX
PN WO200295034-A2.
XX
PD 28-NOV-2002.
XX
PF 23-MAY-2002; 2002WO-CA000758.
XX
PR 23-MAY-2001; 2001US-0292531P.
XX
PA (CYTO-) CYTOCHROMA INC.
XX
PI Wisniewski J, Petkovich PM, Ramshaw H;
XX
DR WPI; 2003-112152/10.
DR N-PSDB; AAD53861.
XX
PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT diagnosing, preventing, or treating disorders with aberrant expression or
PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
XX
PS Example 1; Fig 1; 231pp; English.
XX
CC The invention relates to retinoic acid metabolising cytochrome P450,
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX
SQ Sequence 7052 AA;

Alignment Scores:

Pred. No.:	1.5e-41	Length:	7052
Score:	444.00	Matches:	180
Percent Similarity:	39.9%	Conservative:	47
Best Local Similarity:	31.6%	Mismatches:	139
Query Match:	12.6%	Indels:	205
DB:	6	Gaps:	23

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35266 (1-7052)

```
Qy       1892 TATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATC 1833
Db        104 TyrAlaValLysGluAlaSerTyrLysArgProGlyArgSerPheCysPheValCysPhe 123

Qy       1832 CTTTATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGT 1773
Db        124 ValLeuCysCysPheValLeuAsnArgGlyArgValLeuLeuCysCysProSerTrpSer 143

Qy       1772 GCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTG 1713
Db        144 ThrVal------His-CysSerLeuGlnProProProSerGlyLeuLysLeuAlaSerHi 161

Qy       1712 CCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTG 1653
Db        161 sLeuSerLeuProSerSer---HisTyrArgCysAlaProProGlnProAlaAsnPhe-- 179

Qy       1652 TATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCT 1593
Db        180 -LeuTyr-ValGluThrGlySerHisArgValAlaGlnAlaGlyLeuLysLeuLeuValS 199

Qy       1592 CAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCACTGTG 1533
Db        199 erSer-----------------------------------LeuGlyLeuSerLysCysA 207

Qy       1532 CCCAACTGAGTTATTTTTAATTAATATGAACAAACATCCTCTAATGCAGGTTTCTTGTCT 1473
Db        207 sp------------------------------------------------------- 207

Qy       1472 TAGTTGCAAACATTAAAAAAAAACTGTGTGTGAGTGTGTGTGTGTGTGTGTATTTCTCTG 1413
Db        208 ----------------------------------------------------TyrArgG 210

Qy       1412 GGGAGAGGGTTCATAGCTTTCATCAGATTTGTAAATTTGGCCTTCTAAAGTGGAGCAGGA 1353
```

```
Db        210 lyGluPro-------------------------------------ProTrpAlaArgA 217

Qy       1352 GGATGTATAAGCAAATTAAGCCATAT----------------------------TATT 1323
              ||   :::      |||   |||:::
Db        217 rgProHis---ValIleProProPheGlyGluLeuLeuMetAlaThrGlyPheLeuSerG 236

Qy       1322 CAGTGACTAAAATAGTAACAATGAAAAAACAGTAACAATAAAGGTTTGAGGCAATATGAA 1263
                ||||::::::   |||      |||                :
Db        236 lyAlaLeuLysLeuIleGluMetValAlaGln--------------------PheCysG 249

Qy       1262 AAAGTCATAACAAGGTGATAAGTGAAAAAGCAGAATGC---------------------- 1225
              ::|||   |||      :::      |||      |||
Db        249 luSerGlnAsnHisPheValProPheLysValAsnCysMetValHisLeuLysAlaValL 269

Qy       1224 --AAGATTGTATTTATTCTGCAAAGTTGTCAAGATTGTTTT------------AAAAAAA 1179
                ||||||      ::::::|||      :::      :::             |||   |
Db        269 ysLysIlePheLysValIleGlnGluIleGluValLeuTyrSerTyrLysAsnLysAsnA 289

Qy       1178 GA--------------------AAAAGAAAGTTTATATCTAGGCTGTGCATTCATCAAT 1140
              ||                    :::|||:::   :::|||::::::      |||   |
Db        289 rgGlnLysTrpAspSerTyrProArgArgGluLeuValSerLysValAsnSerHisMetL 309

Qy       1139 TATTTATTGAGCACCTACTATATGTCAGACACTGCTCTAGATCTGGGGATATAGCAATAA 1080
              ||         :::   |||   :::|||      |||      |||   :::
Db        309 eu------AsnAsnLysLeuLysIleArgLysIleSerLeuLeuGlyIleValArgSerL 327

Qy       1079 AGAAAATAGACAGTAA----TCGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACT 1024
                ::::::                |||||   |||   |||   |||      |||||||
Db        327 euGluLeuLysGluGluLeuLysAlaArgArgSerGlySerHisLeu---SerGlnHisP 346

Qy       1023 TTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAAC 964
              ||||||||   |||      ||||||||||||   ||||||   ||||||||||||||||
Db        346 heGlyArgProArgTrpValAspHisLeuArgProGlyVal---AspGlnProGlyGlnH 365

Qy        963 ATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGC 904
              |||||:::||||||||||||   |||   ||| |||||||||      ||||||
Db        365 isGlyGluThrProSerLeuGlnLysThrGln-LysLeuAlaGlyArgGlyGlyMetCys 384

Qy        903 TTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAAC------CCGGGA 850
              |||      |||||||||   |||||||||   :::|||   ||||||      ||||||
Db        385 Leu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnCysLeuAsnLeuArgProGly 403

Qy        849 GGCAGAGGTTGCAGTGAGCTGAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAA 790

Db        404 Ala---------------------------------------------------------- 404

Qy        789 AAAGAAAGGAAGAAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAA 730

Db        404 ---------------------------------------------------------- 404

Qy        729 GATGGAATGTAAAGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACT 670
                                                 |||||||||   ||||||||||||||||
Db        405 --------------------------------ValAlaHisAlaCysAsnProSerThr 413

Qy        669 TTGGGATGCCAAGGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAAC 610
              ||||||   |||    |||   :::    :::   ||||||   |||:::|||   |||
Db        414 LeuGlyGlyGlnAsp-GlyTrpIle-ThrArgSerGlyVal---AspGlnProAspGlnH 432

Qy        609 GTAGTAAGACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACG 550
                 :::|||   ||||||   ||||||      |||   ||||||:::|||      |
Db        432 isGlyGluThrProSerLeuThrLysAsnThrLysAsn---ProGlyMetValAlaCysA 451

Qy        549 CTCGTAGTCCCAGCTACTTGGGAGGCTG------------------------------ 522
              ||   :::   ||||||   ||||||
Db        451 laCysAsnAlaSerTyrSerGlyGlyGlyArgArgThrAlaThrTrpGluValGluValA 471

Qy        521 ----------------------------------------AGGTCAGCAGATCCC---- 507
                                                     :::|||      |||
Db        471 laValSerArgAspHisAlaThrAlaLeuGlnProGlyGlnGlnSerGluThrProSerG 491

Qy        506 ---------------------------TTGAGCTCAGGAGGTTGAGGCTGCAGTGACC 476
                                        |||:::   ||||||   |||||||||:::
Db        491 lnLysLysLysLysLysLysAsnCysLeuAsnProGlyGlyGluGlyCysSerGluP 511

Qy        475 TATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAAC 416
                    ||||||||||||         ||||||||||||:::|||   ||||||||||||
Db        511 roArgLeuHisHisCysThrProValTrpValThrGluArgAspSerValSerLysGlnG 531

Qy        415 AACAAAAAAAAAACCCAA 399
                  |||||:::|||||||||
Db        531 lnGlnGlnLysThrGln 536
```

RESULT 4
AAU31513
ID AAU31513 standard; protein; 215 AA.
XX
AC AAU31513;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2004.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 452; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 215 AA;

Alignment Scores:

Pred. No.:	1.54e-41	Length:	215
Score:	438.50	Matches:	116
Percent Similarity:	59.9%	Conservative:	14
Best Local Similarity:	53.5%	Mismatches:	60
Query Match:	12.4%	Indels:	28
DB:	4	Gaps:	3

```
US-10-607-806-1_COPY_7200_9200 (1-2001) x AAU31513 (1-215)

Qy        439 TCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCT 498
Db          4 SerValThrGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 23

Qy        499 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 558
Db         24 ArgPheArg***PheSerCysLeuSerLeuLeuSerSerTrpAspTyrArgArgProPro 43

Qy        559 CCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAG 618
Db         44 ProArgProAlaAsn-PheCysIlePheSerArgAspMetAlaPheThrThrLeuAlaAr 63

Qy        619 GCTTGTCTTAAACTCCTGGCTTCAAGC-AATCCTCCTACCTTGGCATCCCAAAGTGCTGG 677
Db         63 gLeuValSerAsnSer***ProGlnCysAspProProThrSerAlaSerGlnSerAlaGl 83

Qy        678 GATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAAT 737
Db         83 uIleThrGlyValSerHisArgAlaTrpProValLeuSerSerProGlnProPhe----- 101

Qy        738 AGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGAC 797
Db        102 ----------------------------------------------PhePhePheAspMe 106

Qy        798 AGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC--------------TCACTGCAAC 842
Db        106 tGluSerHisAlaIleThrGlnAlaGlyVal-GlnTrpArgHisLeuGlySerLeuGlnP 126

Qy        843 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 902
Db        126 roProProProMetPheLys***SerSerCysLeuSerLeuLeuSerSerTrpAspTyrA 146

Qy        903 AGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTT---AGCAGAGATGGGGTTTTA 959
Db        146 rgArgProProProArgProAlaAsnPhePhe-ValPheLeu***ArgAspGlyValSer 165

Qy        960 CCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTC 1019
Db        166 ProCys***ProGlyTrpSerArgSerProAspLeuValIleHis-ProProTrpSerPr 185

Qy       1020 CCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1062
Db        185 oLysSerAlaGlyIleThrGlyLeuSerHisCysAlaGlnPro 199

RESULT 5
AAE35264
ID   AAE35264 standard; protein; 7036 AA.
XX
AC   AAE35264;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #1.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7036 AA;

Alignment Scores:
Pred. No.:                3.83e-40      Length:          7036
Score:                    432.00        Matches:         184
Percent Similarity:       40.6%         Conservative:    70
Best Local Similarity:    29.4%         Mismatches:      160
Query Match:              12.3%         Indels:          212
DB:                       6             Gaps:            26

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35264 (1-7036)

Qy       1831 TTTATTTTATTT------ATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCT 1778
Db        123 PheValLeuPheCysValValLeuPheIleGluAlaGlySerCysSerAlaAlaHisPro 142

Qy       1777 GGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTC 1718
Db        143 GlyValGlnPhe------Ile-AlaHisCysSerLeuHisLeuLeuAspSerSerProPr 160

Qy       1717 TCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAAT 1658
Db        160 o---ThrSerAlaSerGlnValAlaSerThrThrGlyValHisHisHisSerGlnLeuIl 179

Qy       1657 TTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCT 1598
Db        179 ePheAsnCysMetSerLysGlnGlyLeu--------------ThrValLeuProArgLe 194

Qy       1597 CACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCA 1538
Db        194 uValSerAsnSerSerGlnAlaLeuAlaSerGlnSerAlaArgIleThrGlyValSerHi 214

Qy       1537 CTGTGCCCAACTGAGTTATTTTTAATTAATATGAACAAACATCCTCTAATGCAGGTTTCT 1478
Db        214 sHisGlyArgProGluAspHisMetLeuPheAspSerHisLeu---------------- 228

Qy       1477 TGTCTTAGTTGCAAACATTAAAAAAAAACTGTGTGTGAGTGTGTGTGTGTGTGTGTATTT 1418
Db        229 ---------------------------------GlySerAspCys------------- 232

Qy       1417 CTCTGGGGAGAGGGTTCATAGCTTTCATCAGATTTGTAAATTTGGCCTT---CTAAAGTG 1361
Db        233 ----------TrpLeuArgGlyPhePheGln-----------GlyLeuAsnLeuArgTr 245

Qy       1360 GAGCAGGAGGATGTATAAGCAAATTAAG--------------CCATATTATTCAGTGAC 1316
Db        245 pLeuHisAsnSerValAsnArgValLysThrIleAspLeuCysProLeuAsnGluLeuTh 265

Qy       1315 TAAAATAGTAACAATGAAAAAACAGTAACAATAAAGGTTTGAGGCAATATGAAAAAGTCA 1256
Db        265 rValTrpTyrIleSerLysLeu-PheLysLysLysPheLeuLysLeuTyr---------- 281

Qy       1255 TAACAAGGTGATAAGTGAAAAAGCAGAATGCAAGATTGTATTTATTCTGCAAAGTTGTCA 1196
```

```
Db        282 -------------------------AsnLysLysArgPheAsnTyrIleLysValIleL 293

Qy       1195 AGATTGTTTTAAAAAAA-----------------GAAAAAGAAAGTTTATATCTA---- 1158
              ||      :::   |||                  ::::::|||||   |||
Db        293 ysThrLysIleAspLysAsnGlyArgIleAlaThrGlnGluGluSerTrpTyrLysValA 313

Qy       1157 -----------------GGCTGTGCATTCATCAATTATTTATTGAGCACCTACTATATGT 1115
                              |||     ||||||
Db        313 rgIleLeuIleCysAspSerCysLysIleIleAsn------------------------ 324

Qy       1114 CAGACACTGCTCTAGATCTGGGGATATAGCAATAAAGAAAATAGACAGTAATCGGCCAGG 1055
                :::   |||:::       |||   :::      |||:::::: :::||||||
Db        325 -----SerLysLeuGluLys-AlaTyrGlyAspAsnHisAsnLysArg-SerArgProAs 342

Qy       1054 CACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTG 995
              |||   |||       |||   |||  ||||||      |||      ||||||
Db        342 pAlaValAlaHisThrCysAsnProSerIleLeuGlyGlyArgGlyGlyTrpIleThr-- 361

Qy        994 AGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATA 935
              |||||||||||||||||||||||||||||||||||||||||   |||    |||
Db        362 -GlyGlnGluPheGluThrSerLeuGlyAsnMetValLysProArgLeu-TyrLysLysH 381

Qy        934 CAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAG 875
              ||||||   |||   :::||||||   ||||||:::   |||      ||||||   |
Db        381 isLysAsn---ProGlyValValValCysAlaCysSerProSerCysLeuGlyGly---G 399

Qy        874 GCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGGTTGAGG 815
              ||   ||||||
Db        399 lyThrArgIleAla---------------------------------------------- 403

Qy        814 TGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCCCTGTTC 755

Db        403 ------------------------------------------------------------ 403

Qy        754 TGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTAGCCGCACATGGT 695
                                                        :::  :::      |
Db        404 ---------------------------------------ThrGlyGlnValArgT 409

Qy        694 GGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGATTGCTTGAAGCCA 635
              |||||||||||||||||||||||||||:::||||||:::||||||  :::      ||
Db        409 rpLeuThrProValIleProAlaLeuTrpGluAlaLysMetGlyGly--SerArgGlyGl 428

Qy        634 GGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAAAAAAATTTTTTT 575
              |||||||:::|||||||||   |||:::|||:::|||   |||      ||||||
Db        428 nGluPheGluThrSerLeuThrAsnMetValLysProHisLeuLeu---LysIleGlnLy 447

Qy        574 AATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCTGAGGTCAG 515
              ||||||:::|||||||||         |||:::|||||||||   |||||||||
Db        447 sIleSerArgAlaTrpTrpArgValProValMetProAlaThrGlnGluAlaGluAlaGl 467

Qy        514 CAGATCCCTTGAGCT------------------------------------------------ 500
                   ||||||
Db        467 yGluLeuLeuGluProGlyArgTrpArgLeuGlnAlaGluIleThrProLeuHisSerAs 487

Qy        500 ------------------------------------------------------------ 500

Db        487 nLeuGlyAsnLysAlaArgLeuArgLeuLysLysLysLysLysLysArgIleAlaIl 507

Qy        499 -CAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACA 441
               :::|||   :::   ||||||:::      |||||||||   |||      |||   ||
Db        507 eArgGluGluLysValAlaValSerArgAspCysThrThrAlaLeuGlnSerGly---Gl 526

Qy        440 GAGCAAGACCCTGTCTCTAAACAACAACAAAAAAAAAACCCAAA---------------- 398
              ||||:::|||||||||   |||||||||   :::|||   |||
Db        526 nSerGluThrLeuSerGlnAsnAsnAsnAsnArgLysHisLysLysGlnIleAsnLysGl 546

Qy        397 -CTGTCACCACTTACAAGC--------------TGTGTGATCTTGGGCAAATAGCTCAG 354
               |||:::   :::||||||                  ::::::
Db        546 uLeuAsnMetValThrSerGluTyrGlySerGluLeuLeuValSerTyrAlaProPheAs 566

Qy        353 CCTTCTGAGCCTCAGTTTTCTT-----------------ATCTGTTAAGTGAGGATAAT 312
                 |||   ::::::|||:::                  :::|||   :::   :::::

Db        566 pTyrLeuCysIleAsnPheIleAsnHisIleAsnSerHisValCys---IleTyrMetLe 585

Qy        311 TGTACCTATCCCATAGTGTTGTTCTGAGGATTAAAGGAGACACTGCCCAGAACCAAGAAA 252
              :   |||                                          |||:::   :::||
Db        585 uPhePro-----------------------------------TrpProLysArgLysLy 593

Qy        251 ATTAACACTAAATCATGGCTGT-----------TGTTACTATTATTTCCCCCCGGCCTA 204
              |           :::|||||||           |||   :::|||      |||   ::
Db        593 sSerLysSerLysTyrGlyCysValIleGlnLeuCysLysPheTyrCysIleProLeuPh 613

Qy        203 C 203
              :
Db        613 e 613
```

RESULT 6
ADC87151
ID ADC87151 standard; protein; 217 AA.
XX
AC ADC87151;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1604.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR N-PSDB; ADC87150.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 2; SEQ ID NO 1604; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ Sequence 217 AA;

Alignment Scores:

Pred. No.:	3.02e-40	Length:	217
Score:	427.50	Matches:	115
Percent Similarity:	52.3%	Conservative:	11
Best Local Similarity:	47.7%	Mismatches:	64
Query Match:	12.1%	Indels:	51
DB:	7	Gaps:	5

US-10-607-806-1_COPY_7200_9200 (1-2001) x ADC87151 (1-217)

```
Qy        299 TGGGAT-----------------AGGTACAATTATCCTCACTTAACAGATAAGAAAACT 340
              ||||||                 |||   |||:::      |||   |||
```

```
Db        17 TrpAspGlnSerSerValGlnPheArgValAsnPheLeuLeuLeuValAsp--------- 33

Qy       341 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 400
                                 |||   :::         |||
Db        34 --------------------PheThrLysMetLysAsnPheCysSerSerGluAsnIle 46

Qy       401 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 460
               |||||||||:::         |||:::   |||   |||   ||||||   |||
Db        47 LeuPhePhePheLeu---------ArgArgSerLeuThrLeuLeuProArgLeuGluCys 63

Qy       461 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 520
               ||||||           ||||||:::|||:::|||       |||   |||       |||
Db        64 SerGlyThrIleSerAlaHisCysAsnLeuHisLeuProGlySerGlyAspSerHisThr 83

Qy       521 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAA 580
               ||||||||||||||||||||    |||       |||   ||||||||||||||||
Db        84 SerAlaSerGlnValAlaGlyIleThrGlyAlaHisArgHisAlaTrpLeuIlePheVal 103

Qy       581 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 640
               |||   :::|||||||||       :::||||||||||||   |||:::||||||   |
Db       104 --PheLeuLeuGluThrGlyPheCysHisValGlyGlnAlaGlyLeuGluLeuLeuThrS 123

Qy       641 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 700
               ||   :::||||||   ||||||||||||||||||||||||||||||||||||||||
Db       123 erGlyAspProProAlaLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHis---- 141

Qy       701 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 760

Db       141 ------------------------------------------------------------ 141

Qy       761 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 820
                        |||      |||||||||||||||||||||||||   :::|||
Db       142 -----ProLysProPhePhePhePhePheGluThrGluSerHisThrValThrTrpA 160

Qy       821 CCTCCGTTCAGC--------------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGA 865
                |||                  ||||||||||||   ||| ||||||||||||||
Db       160 laGlyVal-GlnTrpCysAspLeuGlySerLeuGlnProProPro-ProGlySerSerAs 179

Qy       866 TTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCT 925
               ||||||||||   ||||||   ||||||||||||||      ||||||||||||   ||
Db       179 pSerProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaHisHisHisAlaGlnLe 199

Qy       926 AATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCT 980
               |||||||      |||   ||||||||||||      |||   ||||||||||||
Db       199 uIlePheLeuPhePheArgArgAspGlyValSerLeuCysTrpProGlyTrpSer 217

RESULT 7
AAE35265
ID   AAE35265 standard; protein; 7067 AA.
XX
AC   AAE35265;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #2.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7067 AA;

Alignment Scores:
Pred. No.:                8.57e-39        Length:          7067
Score:                    420.50          Matches:         172
Percent Similarity:       40.0%           Conservative:    51
Best Local Similarity:    30.9%           Mismatches:      148
Query Match:              11.9%           Indels:          188
DB:                       6               Gaps:            20

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAE35265 (1-7067)

Qy       537 CTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGAC 596
               |||||||||:::         |||   ||||||          ||||||   ||||||
Db         2 LeuGlyLeuGlnValProAlaThrMetProGlyPhePheIleTyrPhePheSerArgAsp 21

Qy       597 TGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTAC 656
                   |||   |||   :::||||||   |||      |||   :::      |||      |
Db        22 ArgValSerLeu---TyrProGlyTrpSer--ArgSerHisValIleMetIleHisProP 40

Qy       657 CTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTC 716
               |||||   ||||||||||||||||||||      ||||||:::   |||          |
Db        40 roTrpProProLysValLeuGlyLeuGlnAla---AlaThrValProGlyProAlaAlaL 59

Qy       717 TTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTT 776
               ||::::::   :::   :::            ||||||:::      :::      ::::
Db        59 euPheLeuIleAlaLysLysAspLysCysProSerThrAspAspTrpIleAsnLysValT 79

Qy       777 TCTTCCTT---------------------------------------------------- 784
               ::::::::
Db        79 rpTyrIleHisThrIleTyrTyrTyrLeuSerIleLysArgAsnGluValIleMetHisA 99

Qy       784 ---------------------------------------------------------- 784

Db        99 laThrThrThrLeuLysIleLeuCysLeuSerGluArgSerGlnLeuGlnLysThrArgV 119

Qy       785 ---------------------------------TCTTTTTTGAGACAGAGTCTCACTT 809
                                                  |||   ||||||   |||
Db       119 alPheLeuPheCysLeuPheCysPheValLeuPheCysPheLysArgGlnGlyLeuAlaL 139

Qy       810 CATCACCTCAACCTCCGTTCAGCTCACTGCA-ACCTCTGCCTCCCGG------GTTCAAG 862
                  |||         :::|||||||||   |||:::||||||           |||
Db       139 euLeuProIleLeuGluTyrSerSerLeuLeuThrAlaAlaSerThrPheTrpThrGlnA 159

Qy       863 YGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCAT----G 918
                  |||   |||:::||||||:::   |||   ||||||   |||||||||      :
Db       159 laSerLeuProProGlnProProLys---LeuAlaLeuGlnValCysThrThrThrAlaS 178

Qy       919 CTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGT 978
               :::::||||||   |||         ||:::   |||   ||||||||||||||||||
Db       178 erPheLeuIleValCys---------ArgAsnArgValSerProCysCysProGlyTrpS 195

Qy       979 CTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATA 1038
```

```
Db        195 erGlnThrProSerLeuLys-ProTrpProLeuLysVal---------LeuGlyLeuGln 211

Qy       1039 GGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGA 1098

Db        212 GlyAlaThrMetValGlyGlnLysThrThrCysTyrLeu------------IleProIle 227

Qy       1099 TCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCT 1158

Db        228 TrpGlyValIleAlaAsn---------------------------------------- 233

Qy       1159 AGATATAAACTTTCTTTTTCTTTTTTAAA--------------ACAATCTTGACAACT 1203

Db        234 GlyTyrGlyValSerPheArgGlySerLysIleAspAspGlyCysThrIleLeu------ 251

Qy       1204 TTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTT 1263

Db        252 IleGluSerLysProLeuIleCysAlaLeuMetSerLeuTyrGly---------ThrSer 268

Qy       1264 TCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGA 1323

Db        269 GlnSerCysLeuLys--------------------LysAsnPheSerTyr-IleThrAr 281

Qy       1324 ATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACA 1383

Db        281 gAsnArgGlyLeuIle---------IleLeuLysLeuLysGlnLysThrLysMetValGl 298

Qy       1384 AATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCAC 1443

Db        298 yLeu-----------------------ProLysLysArgAla---------GlyIleLy 307

Qy       1444 ACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCAT-------------- 1489

Db        307 sGlyLysPheSerTyrValIleAlaValLysThrGlnAsnLysAsnLysProIleArgAs 327

Qy       1490 -TAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCC 1548

Db        327 pSerGluIleIleThrArgThrLysArgGlyValLysGlyGlnThrGlnTrpLeuThrPr 347

Qy       1549 TGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAG 1608

Db        347 oValIleProAlaPheTrpGluAlaGluValGlyGlySerLeuGluAlaArgSerLeuAr 367

Qy       1609 ACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGT 1668

Db        367 gProAlaTrpAlaThrTrp---AsnProValSerThrLysAsnThrLysIleSerArgAl 386

Qy       1669 GTAGTGATGCATGCCTGTAGTCCCAGCTACT---------------------------- 1699

Db        386 aTrpTrpTyrValProValValProAlaAlaTrpGluAlaGluAlaGlnGluLeuLeuGl 406

Qy       1699 ---------------------------------------------------------- 1699

Db        406 uProGluAlaArgCysGlyGlySerArgLeuSerGlnHisPheGlyArgProArgTrpVa 426

Qy       1699 ---------------------------------------------------------- 1699

Db        426 lAspHisGluValArgSerLeuArgProAlaProThrTrpAsnProIleSerAsnLysTy 446

Qy       1700 ----------------------------------------------CGGGAGGCTGAGGC 1713

Db        446 rLysLysLeuAlaGlyHisGlyGlyValCysLeuCysGlnLeuLeuArg-ArgLeuArgG 466

Qy       1714 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1773

Db        466 lnGluAsnCysLeuAsnLeuGlyGlyGlyGlyCysSerGluProArgSerArgHisCysT 486

Qy       1774 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAA 1814

Db        486 hrProThrTrpAlaThrLysArgAspSerValSerLysLys 499

RESULT 8
ADA84009
ID   ADA84009 standard; protein; 673 AA.
XX
AC   ADA84009;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human POM77 protein.
XX
KW   human; marker; expressed sequence tag; EST; arabidopsis; tumour;
KW   stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2002103028-A2.
XX
PD   27-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-IB004189.
XX
PR   30-MAY-2001; 2001US-0293999P.
PR   22-OCT-2001; 2001US-0330457P.
PR   19-FEB-2002; 2002US-0357144P.
XX
PA   (BIOM-) BIOMEDICAL CENT.
XX
PI   Baranova AV,  Yankovsky NK,  Kozlov AP,  Lobashev AV,  Krukovskaya LL;
XX
DR   WPI; 2003-175241/17.
DR   N-PSDB; ADA84008.
XX
PT   Determining if a nucleic acid is a marker for a phenotype/cell type of
PT   interest, by global comparison of expressed sequence tags known to be
PT   expressed in the phenotype/cell type with all ESTs expressed in normal
PT   tissue.
XX
PS   Claim 29; Page 392-395; 516pp; English.
XX
CC   The invention relates to a novel method for determining if a nucleic acid
CC   is a marker for a predetermined phenotype/cell type of interest from a
CC   biological species. The method comprises performing a global comparison
CC   of a group of expressed sequence tags (ESTs) known to be expressed in the
CC   phenotype/cell type of interest with all ESTs expressed in normal tissue
CC   in order to identify ESTs that are preferentially expressed in the
CC   phenotype/cell of interest. A method of the invention is useful for
CC   determining whether a nucleic acid is a marker for a predetermined
CC   phenotype or cell type of interest from a biological species, preferably
CC   Arabidopsis or human. The cell type of interest is an abnormal cell such
CC   as a tumour cell, and the predetermined phenotype is a stress-induced
CC   phenotype such as hyperosmotic stress or high salt conditions. A method
CC   of the invention is also useful for determining the progression of colon
CC   cancer in a human, for detecting a tumour cell, and for regulating or
CC   preventing the growth of a tumour cell. An antibody of the invention is
CC   useful for detecting the absence or presence of peptides encoded by
CC   tumour-associated markers. A polypeptide of the invention is useful as an
CC   immunogen for vaccinating an animal. The present sequence represents a
CC   tumour-associated antigen of the invention.
XX
SQ   Sequence 673 AA;

Alignment Scores:
Pred. No.:              1.08e-38        Length:         673
Score:                  416.00          Matches:        131
Percent Similarity:     41.8%           Conservative:   17
Best Local Similarity:  37.0%           Mismatches:     85
Query Match:            11.8%           Indels:         122
DB:                     6               Gaps:           9

US-10-607-806-1_COPY_7200_9200 (1-2001) x ADA84009 (1-673)

Qy        332 AAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGG 391

Db        328 LysLysLysAsnProGlnLys--------------Gln*********PheCysPheTrp 342
```

```
Qy        392 TGACAGTTTGGGTTTTTTTTTGTTGTTGTTTA-------GAGACAGGGTCTTGCTCTGTC 444
              |||   |||||||||       |||         |||   |||   ||||||
Db        343 GlyPhePhePhePhePhePheCysPheValPheValPheGluIleGlyProCysSerAsp 362

Qy        445 ACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTC 504
              |||   |||       ||||||      :::   ||||||||||||||||||||   ||
Db        363 ThrProAlaGlyValGlnTrp-GlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySe 382

Qy        505 AAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGC 564
              |   |||         |||   |||:::|||||||||   |||   :::      |||::
Db        382 rSerAspSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSe 402

Qy        565 CTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGT 624
              :|||||||||       |||   :::||||||     ::::::|||   ||||||
Db        402 rTrpLeuIleTyrVal--PheLeuIleGluThrGlnPheHisHisValAspGlnAlaGly 421

Qy        625 CTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACA 684
              ||||||||||||   ||||||:::   |||::: |||   ||||||||||||||   ||
Db        422 LeuLysLeuLeuThrSerSerAspLeuProSer-TrpSerProLysValLeuGly***Gl 441

Qy        685 GGGGTGAGCCACC----------------------ATGTGCGGCTACTTATTTCTTTA 720
              |     ||||||                      |||:::   |||   ::
Db        441 nAla***AlaThrThrProSer*********PheValPheGlyGlyPhePhePhePh 461

Qy        721 CATTCCA-------------------------------------------------- 727
              :
Db        461 ePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArgLeuGluCysSe 481

Qy        728 ----TCTTTCCAATAGAATGTAAGATCCACA---------GAACAGGGATTACTGCCTAT 774
                  |||           ||||||         :::      |||||||||::
Db        481 rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLe 501

Qy        775 T-----------------------------------------------TTCTTCCTTTC 786
              :                                               |||
Db        501 uSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyrPh 521

Qy        787 TTTTTTGAGACAGAGTCTCACT----------------------TCATCACCTCAA-- 820
                  |||   |||   |||                        |||   ||||||
Db        521 e******ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVa 541

Qy        821 ----------------------------------------------CCTCCG-------- 826
                                                           ||||||
Db        541 lIleCysProProGlyLeuProLysCysTrpAspAspArgArgGluProProHisProAl 561

Qy        826 ------------------------------------------------------------ 826

Db        561 a*********LeuPheLeuGlyValPhePhePhePhePheLeuLeuCysPheCysPhe** 581

Qy        827 ----------------------------------------------TTCAGCTCACTGCA 840
                                                           |||   ||||||||
Db        581 *AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuGl 601

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900
              ||||   ||||||||||||||   ||||||||||||   |||      ||||||:::||
Db        601 nProGlnProProGlyPheLysArgPheSerCysLeuCysLeuProCysSerTrpAspTy 621

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960
              |:::      ||||||   |||||||||    |||||||||:::||||||   |||   |
Db        621 rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP 641

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020
              |||||      ||||||||||||||||||||||||||||   ||||||   ||| ||||||
Db        641 roCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSer 660

Qy       1021 CAAAGTGCTGGAATTATAGGCGTGAGTCAC 1050
              ||||||||||||:::   ||||||||||
Db        661 GlnSerAlaGlyMetThrGlyValSerHis 670
```

RESULT 9
AAO06014
ID AAO06014 standard; protein; 185 AA.
XX

AC AAO06014;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 19906.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI85945.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 19906; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 185 AA;

Alignment Scores:

Pred. No.:	1.42e-38	Length:	185
Score:	413.00	Matches:	98
Percent Similarity:	54.7%	Conservative:	12
Best Local Similarity:	48.8%	Mismatches:	49
Query Match:	11.7%	Indels:	42
DB:	4	Gaps:	3

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAO06014 (1-185)

```
Qy        442 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 501
              :::   ||||||      ||||||||||||      ||||||||||||   ||||||
Db         11 IleAlaGlnAlaGlyValGlnTrpCysAsnLeuSerSerLeuGlnProLeuProProArg 30

Qy        502 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 561
                 |||      :::   ||||||||||||||||||||||||||||          ||||||
Db         31 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisValProPro 50

Qy        562 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 621
              ||||||||||||          |||   ||||||   |||      :::||||||||||||
Db         51 ArgLeuAlaAsnPhePheValPheLeuValGluMetGlyPheLeuHisValGlyGlnAla 70

Qy        622 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACC--------------------TTG 660
```

```
Db         71 SerLeuGluLeuLeuAlaSerGlyAspProProSerGlyAspProAspLeuArg***Leu 90

Qy        661 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 720
Db         91 AlaPheGlnSerLeuGlyIleThrGlyValThrHisHis-AlaGlyGlnIlePheLeuPh 110

Qy        721 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 780
Db        110 eCysLeuGlu---------------------------------------------------- 113

Qy        781 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCG-------------- 826
Db        114 ------------------------ThrGluSerProSerValAlaGlnValGlyValGl 125

Qy        827 ----------TTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 876
Db        125 nTrpCysAsnLeuGlySerLeuGlnProLeuProProGlySerGlu***PheSerCysLe 145

Qy        877 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTA 936
Db        145 uGlyLeuProSerSerTrpAspTyrArgHisAlaLeuThrHisProAlaAsn-PheSerI 165

Qy        937 TTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 995
Db        165 lePheSerArgAspGlyValSerProSerTrpProGlyTrpSerGlnMetProAspLeu 184

RESULT 10
ADK36978
ID   ADK36978 standard; protein; 232 AA.
XX
AC   ADK36978;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID9060.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .232
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 9060; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 232 AA;

Alignment Scores:
Pred. No.:              1.36e-37      Length:          232
Score:                  405.00        Matches:         108
Percent Similarity:     56.1%         Conservative:    11
Best Local Similarity:  50.9%         Mismatches:      52
Query Match:            11.5%         Indels:          42
DB:                     5             Gaps:            5

US-10-607-806-1_COPY_7200_9200 (1-2001) x ADK36978 (1-232)

Qy        400 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 459
Db         52 TrpProPhePhe------PhePheGlnThrGluSerArgSerValThrGlnAlaGlyVal 69

Qy        460 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 519
Db         70 Gln***CysAsnLeuSerSerLeuGlnProLeuProProGlyLeuLysGlnPheSerArg 89

Qy        520 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 579
Db         90 LeuSerLeuProSerSerTrpAspTyrArgHisMetProProArgProAlaSer---Phe 108

Qy        580 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG-GC 638
Db        109 ValPheLeuValGluThr-GlyPheThrMetLeuAlaArgLeuValTrpAsnSer***Pr 128

Qy        639 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC----- 693
Db        128 oSerCysAspProProIleSerAlaSerGlnAsnAlaGlyIleThrGly***SerProSe 148

Qy        694 ----------CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATG 743
Db        148 rLeuAlaProHisHisProArgLeu---------------------------------- 156

Qy        744 TAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTC 803
Db        157 ---------------------------------------PhePhePheGluThrLysSe 163

Qy        804 TCACTTCATCACCTCAACCTCCGTTCAGC--------------TCACTGCAACCTCTGC 848
Db        163 rGlySerValThrGlnAlaGlyVal-GlnTrpHisAsnProThrSerGlnGlnProLeuP 183

Qy        849 CTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTG 908
Db        183 roProGlyPheLysGlnLeuSerArgLeuSerLeuSerSerSerTrpAsp***ArgTyrA 203

Qy        909 CACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGC 968
Db        203 laProSerCysLeuAlaAsn-TyrCysIlePheSerArgAspArgAlaSerProCysTrp 222

Qy        969 CCAGGCTGGTCTCAAACTCCTGACCTCAAG 998
Db        223 SerGlyTrpSer***ThrProAspLeuArg 232

RESULT 11
ADK36936
```

```
ID   ADK36936 standard; protein; 302 AA.
XX
AC   ADK36936;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID9018.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .302
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 9018; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 302 AA;

Alignment Scores:
Pred. No.:              5.84e-37      Length:         302
Score:                  400.00        Matches:        110
Percent Similarity:     53.1%         Conservative:   20
Best Local Similarity:  44.9%         Mismatches:     72
Query Match:            11.4%         Indels:         43
DB:                     5             Gaps:           5

US-10-607-806-1_COPY_7200_9200 (1-2001) x ADK36936 (1-302)

Qy        397 GTTTGGGTTTTTTTTTGTTGTTGT--------------TTAGAGACAGGGTCTTGCTCT 441
              |||      ||||||      |||               |||   ||||||      :::

Db         81 ValCysThrPhePheIleTyrCysProPheLeuPhePheLeuArgThrGlyPheSerAla 100

Qy        442 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 501
              ::::::||||||      ||||||   :::      ||||||||||||   ||||||

Db        101 LeuSerGlnAlaValValGlnTrpArgAspLeuSerSerLeuGlnProProProProGly 120

Qy        502 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 561
              ||||||      :::   ||||||||||||||||||||||||:::         ||||||

Db        121 LeuLysArgPheSerCysLeuSerLeuProSerSerTrpAspHisArgHisValProPro 140

Qy        562 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 621
                  ||||||   |||   :::   |||   |||   |||      ||||||::

Db        141 HisProAlaAsn-PheCysIleLeuMet***ArgTrpGlyPheThrMetPheAlaArgMe 160

Qy        622 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 681
              :      :::               :::|||  |||   |||||||||:::      |||

Db        160 tThrCysAsp--------------LeuLeu--ThrSerAlaSerGlnThrValArgIle 174

Qy        682 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 741
              ||||||||||||||||||   ||||||

Db        175 ThrGlyValSerHisHisThrArgLeu---------------------------------- 183

Qy        742 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTT---------------CTTCCTTTC 786
                                  |||      ||||||               :::   |||

Db        184 --------------------IleTyr***TyrPhe***Ala***SerAspIlePhePhe 196

Qy        787 TTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC-------------- 832
              |||||||||   ||||||      :::           |||

Db        197 PhePheGluLysGluSerCysCysValGlyGlnAlaGlyVal-GlnTrpHisAspLeuGl 216

Qy        833 -TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAG 891
               ||||||||||||   |||   |||   |||   ||||||||||||:::|||   |||||

Db        216 ySerLeuGlnProProProSerGlyLeuLysGlnPheSerCysLeuAsnLeuProSerSe 236

Qy        892 CTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATG 951
              ||||::::::||||||   ||||||          ::: ||||||||||||||||||||

Db        236 rTrpAspHisLysArgThrProProHisProGlySer-PheCysIlePheSerArgAspA 256

Qy        952 GGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCC 1011
                |||   |||       |||||||||       ||||||||||||||||   ||||||   |

Db        256 rgValSerProTrpGlyProGlyTrpPhe***ThrProAspLeuLysArgSerAlaLeuL 276

Qy       1012 TCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTC 1071
              ||   ||||||||||||   :::||||||||||||           |||:::      :

Db        276 euGlyLeuProLysCysCysAspTyrArgArgGluProProArgProAlaGluValThrI 296

Qy       1072 TATTT 1076
              ::|||

Db        296 lePhe 297

RESULT 12
ABG68738
ID   ABG68738 standard; protein; 361 AA.
XX
AC   ABG68738;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human prostate specific protein DEX0293_107.
XX
KW   Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
KW   prostate specific protein; cytostatic; non-cancerous prostate disease;
KW   gene therapy; cancer; immunostimulant; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200255735-A2.
XX
PD   18-JUL-2002.
XX
PF   27-NOV-2001; 2001WO-US044363.
XX
PR   27-NOV-2000; 2000US-0253176P.
```

```
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Recipon H,  Cafferkey R,  Ali S,  Sun Y;
PI   Liu C,  Chen S;
XX
DR   WPI; 2002-557831/59.
DR   N-PSDB; ABK97631.
XX
PT   New prostate specific genes, useful for treating or diagnosing cancer, or
PT   useful as vaccines for treating cancer, particularly prostate cancer, in
PT   a patient.
XX
PS   Claim 11; Page 207-208; 212pp; English.
XX
CC   The invention relates to a new isolated prostate-specific nucleic acid
CC   (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
CC   ABK97642 which encode prostate specific proteins appearing as ABG68701-
CC   ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
CC   homology with a PSNA. Also included are a method of determining the
CC   presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
CC   comprising the vector, producing the polypeptide encoded by the PSNA, a
CC   method of determining the presence of a PSP in a sample, diagnosing and
CC   monitoring the presence and metastases of prostate cancer in a patient, a
CC   kit for detecting a risk of cancer or presence of cancer in a patient
CC   (the kit comprising a means for determining the presence of the PSNA or
CC   PSP in a sample of a patient) and a vaccine comprising the polypeptide or
CC   the nucleic acid encoding the polypeptide. The PSNA, PSP and anti-PSP
CC   antibody are useful for diagnosing and treating cancer in a patient (e.g.
CC   by gene therapy). The nucleic acid molecule and polypeptide are also
CC   useful as vaccines for treating cancer, particularly prostate cancer and
CC   non-cancerous prostate diseases. The present sequence is a PSP of the
CC   invention
XX
SQ   Sequence 361 AA;

Alignment Scores:
Pred. No.:              6.3e-37         Length:         361
Score:                  400.00          Matches:        105
Percent Similarity:     55.8%           Conservative:   11
Best Local Similarity:  50.5%           Mismatches:     55
Query Match:            11.4%           Indels:         38
DB:                     5               Gaps:           5

US-10-607-806-1_COPY_7200_9200 (1-2001) x ABG68738 (1-361)

Qy        439 TCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCT 498
              ||||||   ||||||       ||||||||||||      ||||||||||||   ||||||
Db        172 SerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 191

Qy        499 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 558
                  :::      :::   |||||||||   ||||||||||||:::      ||||||
Db        192 ArgPheArgArgPheSerCysLeuSerLeuLeuSerSerTrpAspHisArgHisAlaPro 211

Qy        559 CCACGCCTGGCTAATTAAAAA------AATTTTTTTGTAGAGACTGGGTCTTACTACGTT 612
              |||   |||||||||               |||   |||:::          |||   ||
Db        212 ProCysLeuAlaAsnPheLeuPhePheLysPheLeuValAspGln-SerPheThrMetLe 231

Qy        613 GGCCAGGCTTGTCTTAAACTCCTG-GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAG 671
              |||||||||||||||||||||      |||   :::   |||      ||||||||||
Db        231 uAlaArgLeuValLeuAsnSerAlaProSerGlyAspLeuProAlaProAlaSerGlnSe 251

Qy        672 TGCTGGGATTACAGGGGTGAGCCAC---------------------------------- 696
              ||||||||||||||||||   |||
Db        251 rAlaGlyIleThrGlyValArgHisCysThrTrpProLysSerPheLysPheAlaAspSe 271

Qy        697 -CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAG 755
               |||:::   |||      :::   |||      |||
Db        271 rHisIleGlyLeu------AlaPheHisPheAlaPhe---------------------- 281

Qy        756 AACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCAC 815
                      :::|||      |||||||||         ||||||   |||
Db        282 ---------------PhePhePhePhePhePheAlaValAlaSerHisProIleAl 296

Qy        816 CTCAACCTCCGTTCAGC--------------TCACTGCAACCTCTGCCTCCCGGGTTCA 860
                  |||                   ||||||||||||   ||||||||||||
Db        296 aGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnProProProProGlyPheL 316

Qy        861 AGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCT 920
              ||   |||   ||||||||||||      ||||||:::|||:::||||||||||||
Db        316 ysGlnPheLeuCysLeuSerLeuProGlySerTrpAspTyrArgArgAlaProProArgG 336

Qy        921 TGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCT 980
                |||||| ||||||||||||||||||||||||||||   ||||||      |||||||||
Db        336 lnAlaAsn-PheCysIlePheSerArgAspGlyValSerProCysTrpThrGlyTrpSer 355

Qy        981 CAAACTCCTGACCTCAAG 998
              ||||||||||||||:::
Db        356 GlnThrProAspLeuArg 361

RESULT 13
ABJ01057
ID   ABJ01057 standard; protein; 449 AA.
XX
AC   ABJ01057;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Ovary cell-specific amino acid sequence 3.
XX
KW   Ovary cell; neoplastic ovary cell; ovary specific nucleic acid;
KW   ovary specific protein; ovarian cancer; breast cancer; vaccine;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200238606-A2.
XX
PD   16-MAY-2002.
XX
PF   07-NOV-2001; 2001WO-US046459.
XX
PR   08-NOV-2000; 2000US-0246640P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y,  Recipon H,  Salceda S,  Liu C;
XX
DR   WPI; 2002-519297/55.
XX
PT   Polypeptide and polynucleotides present in normal and neoplastic ovary
PT   cells, useful for identifying, monitoring, staging, diagnosing,
PT   preventing and treating ovarian cancer, and non-cancerous disease states
PT   in the ovary.
XX
PS   Claim 11; Page 207-209; 247pp; English.
XX
CC   The invention comprises amino acid and DNA sequences which are present in
CC   normal and neoplastic ovary cells. The DNA and protein sequences of the
CC   invention are useful for determining the presence of an ovary specific
CC   nucleic acid or an ovary specific protein in a sample. The DNA and
CC   protein sequences of the invention are useful for diagnosing and
CC   monitoring the presence and metastasis of ovarian cancer and breast
CC   cancer. Amino acids ABJ01055 - ABJ01155 represent the ovary cell specific
CC   protein sequences of the invention
XX
SQ   Sequence 449 AA;

Alignment Scores:
Pred. No.:              6.9e-37         Length:         449
Score:                  400.00          Matches:        103
Percent Similarity:     58.1%           Conservative:   26
Best Local Similarity:  46.4%           Mismatches:     62
Query Match:            11.4%           Indels:         31
DB:                     5               Gaps:           7
```

```
US-10-607-806-1_COPY_7200_9200 (1-2001) x ABJ01057 (1-449)

Qy        388 GTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACC 447
              ::::::         ::::::|||       |||              ||||||
Db        243 LeuIleAsnAlaLeuIleTyrPhe---------LeuArgGlnSerLeuArgSerValAla 259

Qy        448 CAGGCATGAGCACAGTGGTGCAACCATAGG---------TCACTGCAGCCTCAACCTCCT 498
              ||||||      |||||||||:::              ||||||||||||   ||||||
Db        260 GlnAlaGlyValGlnTrpCysSerGlyAlaAspLeuGlySerLeuGlnProLeuProPro 279

Qy        499 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 558
                  |||          |||||||||   ||||||||||||||         |||
Db        280 GlyPheLysAlaProCysLeuSerLeuLeuSerSerTrpAspTyrArgSerLeuPro 299

Qy        559 CCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAG 618
              |||      ||||||        |||   :::|||||||||   :::::::::   |||
Db        300 ProCysProAlaAsnPhePheValPheLeuIleGluThrGlyPheHisHisIleSerGln 319

Qy        619 GCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGG 678
                  :::   :::   |||      :::||||||      ||||||||||||||||||
Db        320 ------IleSerIleSerAlaProCysAspProProAlaSerAlaSerGlnSerAlaGly 337

Qy        679 ATTACAGGGGTGAGCCACCATGTGCG---------GCTACTTATTTCTTTACATTCCATC 729
              |||||||||:::||||||  ||||||              |||:::   :::   ::::
Db        338 IleThrGlyMetSerHis--CysAlaGlnProAspValTyrTyrTyrValSerGlyTyrI 357

Qy        730 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 789
              ::                  |||:::|||   |||    :::|||      ||||||
Db        357 leGlyLys--------------GlnAspArgCysTyr-LeuPhePhePhePhePhePhe 371

Qy        790 TTTGAGACAGAGTCTCACTTCATCACC--------------------TCAACCTCCGTT 828
              ||||||||||||||      :::                     |||   ::::::
Db        372 PheGluThrGluSerArgThrValAlaGlnAlaGlyArgLeuGluArgSerGlyAlaIle 391

Qy        829 CAGC---------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 879
               |||         ||||||||||||||||||||||||   |||   ||||||||||||||
Db        392 -SerThrArgArgSerLeuGlnProLeuProProGlyLeuLysArgPheSerCysLeuSe 411

Qy        880 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 939
              |||||||||||||||:::|||:::      ||||||   ||||||::: ||||||   |
Db        411 rLeuLeuSerSerTrpAspTyrArgCysThrProProArgLeuAlaHis-PheCysThrP 431

Qy        940 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 995
              |||||||||||||||||   ||||||      |||||||||   :::|||||||||
Db        431 heSerArgAspGlyValSerProCysTrpSerGlyTrpSerLeuSerProAspLeu 449

RESULT 14
AAU30391
ID   AAU30391 standard; protein; 304 AA.
XX
AC   AAU30391;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #882.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 287; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 304 AA;

Alignment Scores:
Pred. No.:              1.15e-36      Length:          304
Score:                  397.50        Matches:         123
Percent Similarity:     49.5%         Conservative:    22
Best Local Similarity:  42.0%         Mismatches:      62
Query Match:            11.3%         Indels:          87
DB:                     4             Gaps:            7

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAU30391 (1-304)

Qy        434 CTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAAC 493
              ::::::|||||   |||   |||   :::||||||   :::   ||||||:::|||:::
Db          3 ValSerLeuSerLeuArgLeuGluCysAsnGlyAlaIleLeuProHisCysAsnLeuHis 22

Qy        494 CTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTG 553
              |||    |||   |||          ||||||||||||||||||||   |||   |||
Db         23 LeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThrGlyVal 42

Qy        554 CACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTG 613
                ||||||||      |||      :::     ||||||||||   ::::::||||
Db         43 CysHisHisAlaProProIlePheValLeu--LeuValGluThrGlyPheHisHisValG 62

Qy        614 GCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTG 673
              |||||||   |||:::||||||   ||||||:::   |||      ||||||||||||
Db         62 lyGlnAlaGlyLeuGluLeuLeuThrSerSerAspLeuProAlaTrpAlaSerGlnSerA 82

Qy        674 CTGGGATTACAGGGGTGAGCCAC--------------------------------CATG 700
              |||||||||||||   ||||||                                  |||
Db         82 laGlyIleThrGlyGlySerHisLeuCysLeuAlaAsnValLysTyrPheLysSerHisP 102

Qy        701 TGCGGCTA----------------------------------CTTATTTCTTTACAT- 723
                ::::::                                 |||:::||||||
Db        102 heLysIlePheValIleAspGluThrTrpPheGlnHisThr***LeuLeuSerLeuSerS 122

Qy        724 -------------------TCCATCTTTCCA-----------------ATAGAATGTA 745
                                  ||||||   |||               :::   |||
Db        122 erCysAsp***Val***ThrSerIle***ProGlnValLeuAsp***AsnValLeuCysP 142

Qy        746 AGATCCACAGAACAGGGATTACTGCCT------------------------------ 772
                  |||   |||||||||
Db        142 ro-AspValSerGln***LeuLeuProTrpLeuIleLysIlePheLeuLysTyrThrLeu 161

Qy        772 ------------------------------------------------------------ 772
```

```
Db        162 AspCysTrp***LysArgGlnAlaTrpLeuTyrTrpTyr***PheGlyThrGluLeuPhe 181

Qy        773 --------------------------ATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTC 805
                                         ||||||      |||||||||:::||||||
Db        182 GlyCysPro***GlyArgGlnThrLeuCysPhePhe------PheLeuArgArgSerLeu 199

Qy        806 ACTTCATCACCTCAACCT---------------CCGTTCAGCTCACTGCAACCTCTGCCT 850
                |||       |||                 ||||||||||||||||||
Db        200 AspSerValAlaGlnAlaGlyValGln***CysAspLeuGlySerLeuGlnProLeuPro 219

Qy        851 CCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCA 910
              |||   ||||||||||||||||||||||||      :::||||||:::|||:::   |||
Db        220 ProArgPheLys***PheSerCysLeuSerPheProAsnSerTrpAspTyrArgCysAla 239

Qy        911 CCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCC 970
              ||||||      |||||||||      |||   :::      ||||||   ||||||
Db        240 ProProLeuProAlaAsnPheCysIlePheSerArgAspMetGlyPheAlaMetLeuVal 259

Qy        971 AGGCTGGTCTCAAA-CTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 1029
              ||||||||||||   ||||||||||||   |||   |||||||||   ||||||||||||
Db        260 ArgLeuValSerGluLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSerAla 279

Qy       1030 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1062
              ||||||   |||:::|||:::   |||||||||
Db        280 GlyIleThrGlyMetSerTyrArgAlaTrpPro 290
```

RESULT 15
AAB01399

```
ID   AAB01399 standard; protein; 375 AA.
XX
AC   AAB01399;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Neuron-associated protein.
XX
KW   Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW   ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW   Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
KW   Parkinson's disease; demyelinating disease; meningitis; prion disease;
KW   kuru; Creutzfeldt-Jakob disease; neurofibromatosis; cerebral palsy;
KW   muscular dystrophy; central nervous system; CNS;
KW   peripheral nervous system; PNS; myopathy; schizophrenia;
KW   actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW   myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW   autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW   AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
KW   ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW   Werner syndrome, trauma; human.
XX
OS   Homo sapiens.
XX
PN   WO200034477-A2.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US030408.
XX
PR   11-DEC-1998;   98US-00210083.
PR   09-FEB-1999;   99US-0119365P.
PR   16-MAR-1999;   99US-0124687P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Hillman JL,  Lal P,  Au-Young J;
PI   Yang J,  Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-423423/36.
XX
PT   New human neuron-associated proteins and polynucleotides encoding them,
PT   useful for diagnosis, treatment and prevention of cell proliferative
PT   disorders including cancer, neuronal and neurological disorders.
XX
PS   Disclosure; Page 143-144; 145pp; English.
XX
CC   Human neuron-associated proteins (NEUAP) can be used for for treating or
CC   preventing a disorder associated with decreased expression or activity of
CC   NEUAP. Antagonists of NEUAP are useful for treating or preventing
CC   disorder associated with increased expression or activity of NEUAP. NEUAP
CC   or their fragments or derivatives are useful for treating neurological
CC   disorder such as epilepsy, ischemic cerebrovascular disease, stroke,
CC   cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's
CC   disease, dementia and Parkinson's disease. NEUAPs are also useful for
CC   treating other demyelinating diseases, bacterial and viral meningitis,
CC   prion diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC   metabolic diseases of the nervous system, neurofibromatosis, other
CC   developmental disorders of the central nervous system, cerebral palsy,
CC   neuroskeletal disorders, autonomic nervous system disorders, cranial
CC   nerve disorders, spinal cord diseases, muscular dystrophy and other
CC   neuromuscular disorders, peripheral nervous system disorders, inherited,
CC   metabolic, endocrine, and toxic myopathies, mental disorders including
CC   mood, anxiety and schizophrenic disorders, a cell proliferative disorder
CC   such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
CC   cirrhosis, hepatitis, mixed connective tissue disease (MCTD),
CC   myelofibrosis, paroxysmal nocturnal haemoglobinuria, cancers of the
CC   adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, and an
CC   autoimmune/inflammatory disorder such as acquired immunodeficiency
CC   syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
CC   allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, Werner
CC   syndrome, complications of cancer, hemodialysis, and extracorporeal
CC   circulation, viral, bacterial, fungal parasitic, protozoal, and
CC   helminthic infections, and trauma. This protein was designated g3002527
XX
SQ   Sequence 375 AA;
```

Alignment Scores:

Pred. No.:	1.26e-36	Length:	375
Score:	397.50	Matches:	105
Percent Similarity:	56.0%	Conservative:	12
Best Local Similarity:	50.2%	Mismatches:	60
Query Match:	11.3%	Indels:	32
DB:	3	Gaps:	4

US-10-607-806-1_COPY_7200_9200 (1-2001) x AAB01399 (1-375)

```
Qy        397 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 456
              :::::::::|||             |||                |||||||||||||||
Db        194 IlePheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213

Qy        457 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 516
                ||||||   |||      ||||||||||||   ||||||      |||      :::
Db        214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSer 233

Qy        517 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 576
                  ||||||   ||||||||||||||||   |||   ||||||||||||||||||
Db        234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252

Qy        577 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 636
                  |||   |||||| :::|||   |||      |||||||||:::|||   |||
Db        253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuIleLeuIleSer-- 271

Qy        637 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 696
                    :::   |||      ||||||||||||||||||||||||||||||||||||
Db        272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 291

Qy        697 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 756
              |||   ||||||:::   :::                              |||
Db        292 HisAlaArgLeuIlePheAsnPhe--------------------Cys------------ 300

Qy        757 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 816
                                                |||||   |||||||||   :::|||
Db        301 ----------------------------LeuPheGluMetGluSerHisSerValThr 310

Qy        817 TCAA--------------CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 861
                |||               |||        ||||||||||||||||||||||||   ||
```

```
Db        311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLy 330

Qy        862 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 921
              |   ||||||||||||||||||   |||||||||:::|||         ||||||
Db        330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisPr 350

Qy        922 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 981
               |||||| ||||||||||||   |||   ||||||   |||         ||||||||||
Db        350 oAlaAsn-PheCysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerG 370

Qy        982 AAACTCCTGACCTCAAG 998
              ||||||||||||||:::
Db        370 lnThrProAspLeuArg 375
```

Search completed: February 8, 2006, 10:00:48
Job time : 387.279 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47 ; Search time 19.8358 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 1253
Sequence: 1 gtcgtgtcactgctgtccag.....tgcagcctcaacctcctgag 701

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092016_27646/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_605_@runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	260	20.8	148	5	AAU78668	Aau78668 Human Pho
2	258	20.6	148	2	AAR26410	Aar26410 Human pho
3	258	20.6	148	2	AAR63054	Aar63054 Human PLA
4	258	20.6	148	5	AAU78667	Aau78667 Human Pho
5	258	20.6	148	6	ABR44232	Abr44232 Human sec
6	258	20.6	148	6	ABU63122	Abu63122 Human gro
7	258	20.6	148	8	ADI35083	Adi35083 Human PLA
8	258	20.6	148	8	ADJ09984	Adj09984 Human pho
9	258	20.6	156	3	AAB54059	Aab54059 Human pan
10	258	20.6	194	8	ABM83696	Abm83696 Human dia
11	258	20.6	217	4	ABG11912	Abg11912 Novel hum
12	244	19.5	145	8	ADO28599	Ado28599 Human PA2
13	229	18.3	79	3	AAB54421	Aab54421 Human pan
14	222	17.7	146	5	ABB08153	Abb08153 Murine PL
15	216	17.2	124	2	AAR63059	Aar63059 Rat PLA2
16	216	17.2	146	2	AAR63052	Aar63052 HPLA2-8.
17	209	16.7	145	2	AAR04821	Aar04821 D-helix d
18	209	16.7	145	3	AAY88294	Aay88294 Bovine PL
19	204	16.3	146	9	ADY72838	Ady72838 Pig phosp
20	204	16.3	146	9	ADY72839	Ady72839 Pig phosp
21	204	16.3	146	9	ADY72840	Ady72840 Pig phosp
22	200	16.0	145	1	AAP94659	Aap94659 Amino aci
23	195	15.6	146	9	ADY72845	Ady72845 Pig phosp
24	195	15.6	146	9	ADY72844	Ady72844 Pig phosp
25	195	15.6	146	9	ADY72841	Ady72841 Pig phosp
26	195	15.6	146	9	ADY72843	Ady72843 Pig phosp
27	195	15.6	146	9	ADY72842	Ady72842 Pig phosp
28	143	11.4	133	6	ADB19008	Adb19008 Gamma-tai
c 29	130.5	10.4	110	5	ABB06221	Abb06221 Phosphoen
30	128	10.2	25	9	ADY38233	Ady38233 Human CPP
31	125.5	10.0	74	4	AAO11433	Aao11433 Human pol
c 32	125	10.0	7052	6	AAE35266	Aae35266 Human P45
c 33	124.5	9.9	80	5	AAM47619	Aam47619 Human chl
c 34	123	9.8	100	4	AAO04010	Aao04010 Human pol
35	122	9.7	118	1	AAP90076	Aap90076 Phospholi
36	118	9.4	150	4	AAO01959	Aao01959 Human pol
37	117	9.3	53	4	AAO11877	Aao11877 Human pol
38	117	9.3	53	5	ABP03853	Abp03853 Human ORF
c 39	116	9.3	115	4	AAM83664	Aam83664 Human imm
40	114.5	9.1	125	4	AAU30406	Aau30406 Novel hum
41	114.5	9.1	127	4	AAU21236	Aau21236 Human nov
42	114	9.1	61	4	AAO12650	Aao12650 Human pol
43	114	9.1	118	3	AAY67383	Aay67383 Acanthin
c 44	112.5	9.0	475	8	ADX94147	Adx94147 Plant ful
c 45	112	8.9	122	4	AAO12468	Aao12468 Human pol

ALIGNMENTS

RESULT 1
AAU78668
ID AAU78668 standard; protein; 148 AA.
XX
AC AAU78668;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human Phospholipase A2, groupIB (PLA2G1B) variant.
XX
KW Human; SNP; single nucleotide polymorphism; pancreatitis; enzyme;
KW pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 122
FT /note= "Wild-type Arg substituted by His"
XX
PN WO200212562-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US024663.
XX
PR 04-AUG-2000; 2000US-0223179P.
XX
PA (GENA-) GENAISSANCE PHARM INC.
XX
PI Kazemi A, Kliem SE, Koshy B;
XX
DR WPI; 2002-303982/34.

```
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 30; Page; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The PLA2G1B gene is located on chromosome 12q23-q24.1.
CC   The present sequence is the PLA2G1B protein variant. Note: The present
CC   sequence is not shown in the specification but was created by the indexer
CC   using the PLA2G1B sequence appearing as AAU78667 and the information in
CC   figure 3
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              5.63e-23      Length:         148
Score:                  260.00        Matches:        60
Percent Similarity:     54.5%         Conservative:   13
Best Local Similarity:  44.8%         Mismatches:     26
Query Match:            20.8%         Indels:         35
DB:                     5             Gaps:           4

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAU78668 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                 :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                ||:::                                    |||   :::||||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG-----------------TTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCT 400
                |:::                    ||||||   :::      :::         :::
Db        108 rLysAsnLysGluCysGluAlaPheIleCysAsnCysAspHisAsnAlaAlaIleCysPh 128

Qy        401 CAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGAT 440
                |||:::        |||   ::::::   :::   |||
Db        128 eSerLysAlaProTyrAsnLysAlaHisLysAsnLeuAsp 141

RESULT 2
AAR26410
ID   AAR26410 standard; protein; 148 AA.
XX
AC   AAR26410;
XX
DT   04-MAR-1993  (first entry)
XX
DE   Human phospholipase A2 type I.
XX
KW   PLA2; rheumatism; osteoporosis; inflammation; immunodeficiency;
KW   immunostimulant; vascular smooth muscle cell; synoviocyte; chondrocyte;
KW   lymphocute.
XX
OS   Homo sapiens.
XX
PN   JP04253921-A.
XX
PD   09-SEP-1992.
XX
PF   06-FEB-1991;   91JP-00036789.
XX
PR   06-FEB-1991;   91JP-00036789.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 1992-352746/43.
XX
PT   Cell growth agents contg. group I phospholipase A2 - used for treating
PT   gastric mucous membrane injury, rheumatism, osteoporosis, inflammation,
PT   immunodeficiency or vascular injury.
XX
PS   Claim 7; Page 2, 9-10; 10pp; Japanese.
XX
CC   Group I type phospholipase A2 can specifically bind ca. 90% to cells such
CC   as vascular smooth muscle cell, synoviocyte, chondrocyte, gastric mucous
CC   membrane cell, and spleen-originated lymphocyte. It is used in remedies
CC   for gastric mucous membrane injury, rheumatism, osteoporosis,
CC   inflammation, immunodeficiency, or vascular injury. Since the agent can
CC   increase the growth of spleen-originated lymphocyte, it is utilizable as
CC   immunostimulant. Pref. 1 nM or higher concns. of the phospholipase is
CC   contacted with cells. Daily dose will pref. be 1mg-100mg given orally or
CC   by injection
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              1e-22         Length:         148
Score:                  258.00        Matches:        55
Percent Similarity:     59.8%         Conservative:   6
Best Local Similarity:  53.9%         Mismatches:     12
Query Match:            20.6%         Indels:         29
DB:                     2             Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAR26410 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                 :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                ||:::                                    |||   :::||||||||
```

```
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109

RESULT 3
AAR63054
ID   AAR63054 standard; protein; 148 AA.
XX
AC   AAR63054;
XX
DT   25-MAR-2003  (revised)
DT   15-AUG-1995  (first entry)
XX
DE   Human PLA2 type I.
XX
KW   HPLA2-10; phospholipase A2; PLA2.
XX
OS   Homo sapiens.
XX
PN   WO9502328-A1.
XX
PD   26-JAN-1995.
XX
PF   15-JUL-1994;   94WO-US007926.
XX
PR   15-JUL-1993;   93US-00091941.
PR   26-JUL-1993;   93US-00097354.
XX
PA   (INDV ) UNIV INDIANA FOUND.
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tischfield JA,  Seilhamer JJ;
XX
DR   WPI; 1995-067096/09.
XX
PT   Novel type III and IV low mol. wt. phospholipase A2 enzymes - from humans
PT   and rats, also nucleic acid sequences useful, e.g. for recombinant prodn.
PT   of enzymes, research into Batten's disease, etc.
XX
PS   Disclosure; Fig 14; 160pp; English.
XX
CC   A human cDNA (AAQ81138) expressing a novel PLA2, HPLA2-10, was derived
CC   from human brain RNA by RACE-PCR. HPLA2-10 (AAR63046) was characterized
CC   as a novel type of PLA2, type IV, on the basis of its Cys content in
CC   comparison with human PLA2 types I (AAR63054) and II (AAR63055) (Updated
CC   on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              1e-22          Length:        148
Score:                  258.00         Matches:       55
Percent Similarity:     59.8%          Conservative:  6
Best Local Similarity:  53.9%          Mismatches:    12
Query Match:            20.6%          Indels:        29
DB:                     2              Gaps:          3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAR63054 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                 :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                          |||   :::||||||||
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109

RESULT 4
AAU78667
ID   AAU78667 standard; protein; 148 AA.
XX
AC   AAU78667;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B).
XX
KW   Human; SNP; single nucleotide polymorphism; pancreatitis; enzyme;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   N-PSDB; ABK47376, ABK47377.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 30; Fig 3; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
```

```
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The PLA2G1B gene is located on chromosome 12q23-q24.1.
CC   The present sequence is the PLA2G1B protein
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              1e-22          Length:         148
Score:                  258.00         Matches:        55
Percent Similarity:     59.8%          Conservative:   6
Best Local Similarity:  53.9%          Mismatches:     12
Query Match:            20.6%          Indels:         29
DB:                     5              Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAU78667 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
                 |||||||||   |||   |||:::                               |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                  :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                 |||:::                                 |||   :::||||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
                 |   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                 |:::
Db        108 rLys 109

RESULT 5
ABR44232
ID   ABR44232 standard; protein; 148 AA.
XX
AC   ABR44232;
XX
DT   18-AUG-2003  (first entry)
XX
DE   Human secreted group IB phospholipase A2 (sPLA2) enzyme.
XX
KW   Group IIF secreted phospholipase A2; sPLA2; phosphatidylglycerol; human;
KW   phosphatidylcholine; antibacterial; virucide; cytostatic; vasotropic;
KW   antiinflammatory; vulnerary; cardiant; chromosome 1p35; transgenic;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO2003033689-A1.
XX
PD   24-APR-2003.
XX
PF   12-OCT-2001; 2001WO-IB002407.
XX
PR   12-OCT-2001; 2001WO-IB002407.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Lazdunski M,  Lambeau G,  Valentin E;
XX
DR   WPI; 2003-403216/38.
XX
PT   Novel mammalian secreted group IIF secreted phospholipase A2, useful for
PT   preventing and treating bacterial and viral infections, and cancers.
XX
PS   Disclosure; Fig 1; 33pp; English.
XX
CC   The invention relates to a mammalian secreted group IIF secreted
CC   phospholipase A2 (sPLA2) (I), where the enzyme is Ca2+ dependent,
CC   maximally active at pH 7-8 and hydrolyzes phosphatidylglycerol versus
CC   phosphatidylcholine with a 15-fold preference. A pharmaceutical
CC   composition comprising (I) is useful for treating or preventing viral and
CC   bacterial infections, and cancers. A pharmaceutical composition
CC   containing compounds capable of inhibiting catalytic activity of (I),
CC   biologically active compounds that bind sPLA_2 receptors, or a compound
CC   that modulates cell roliferation, cell migration, cell contraction or
CC   apoptosis is useful for treating disease states or disorders involving
CC   group IIF sPLA2, such as inflammatory disease, cancers, cardiac and brain
CC   ischaemia, acute lung injury, acute respiratory distress syndrome or
CC   Crohn's disease. Specific antibodies are useful for searching new
CC   secreted mammalian group IIF sPLA2 or the homologues of the enzyme in
CC   other mammals. The encoding polynucleotides and vectors are useful for
CC   transforming animals and establishing a line of transgenic animals.
CC   Sequences ABR44232-238 represent various human secreted sPLA2 enzymes
CC   used in alignment studies with the GIIF sPLA2 enzyme
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              1e-22          Length:         148
Score:                  258.00         Matches:        55
Percent Similarity:     59.8%          Conservative:   6
Best Local Similarity:  53.9%          Mismatches:     12
Query Match:            20.6%          Indels:         29
DB:                     6              Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABR44232 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
                 |||||||||   |||   |||:::                               |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                  :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                 |||:::                                 |||   :::||||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
                 |   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                 |:::
Db        108 rLys 109

RESULT 6
ABU63122
ID   ABU63122 standard; protein; 148 AA.
XX
AC   ABU63122;
```

```
XX
DT   25-SEP-2003  (first entry)
XX
DE   Human group IIB secreted phospholipase A2.
XX
KW   Human; group IIB secreted phospholipase A2; virucide; antibacterial;
KW   cytostatic; antiinflammatory; vasotropic; cerebroprotective; sPLA2;
KW   phosphatidylglycerol hydrolysis; phosphatidylcholine hydrolysis;
KW   viral infection; bacterial infection; cancer; inflammatory disease;
KW   cardiac ischaemia; brain ischaemia; acute lung injury;
KW   acute respiratory distress syndrome; Crohn's disease; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2003073087-A1.
XX
PD   17-APR-2003.
XX
PF   11-OCT-2001; 2001US-00975456.
XX
PR   11-OCT-2000; 2000US-0239491P.
XX
PA   (LAZD/) LAZDUNSKI M.
PA   (LAMB/) LAMBEAU G.
PA   (VALE/) VALENTIN E.
XX
PI   Lazdunski M,  Lambeau G,  Valentin E;
XX
DR   WPI; 2003-567302/53.
XX
PT   New mammalian secreted group IIF phospholipase A2 or nucleic acid, useful
PT   for treating or preventing viral or bacterial infections, or cancers, or
PT   screening inhibitors of the enzyme for treating e.g. inflammatory
PT   diseases or ischemia.
XX
PS   Disclosure; Fig 1; 16pp; English.
XX
CC   The invention describes a mammalian secreted group IIF phospholipase A2
CC   (sPLA2), which is Ca2+-dependent, maximally active at pH of about 7-8,
CC   and hydrolyses phosphatidylglycerol versus phosphatidylcholine with about
CC   a 15-fold preference. The mammalian secreted group IIF sPLA2 protein or
CC   nucleic acid, or a pharmaceutical composition is useful for treating
CC   and/or preventing viral infections, bacterial infections, or cancers. The
CC   inhibitors of sPLA2 or a composition comprising sPLA2 inhibitors is
CC   useful for treating disease states or disorders involving group IIF
CC   sPLA2, e.g. inflammatory diseases, cancer, cardiac and brain ischaemia,
CC   acute lung injury, acute respiratory distress syndrome, or Crohn's
CC   disease. The enzyme is also useful for screening various chemical
CC   compounds for treating these diseases. This is the amino acid sequence of
CC   human group IIB phospholipase A2 used to determine a consensus sequence
CC   for human sPLA2s
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:                1e-22          Length:          148
Score:                    258.00         Matches:         55
Percent Similarity:       59.8%          Conservative:    6
Best Local Similarity:    53.9%          Mismatches:      12
Query Match:              20.6%          Indels:          29
DB:                       6              Gaps:            3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABU63122 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                  :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                ||:::                                  |||   :::||||||||
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                |:::
Db        108 rLys 109

RESULT 7
ADI35083
ID   ADI35083 standard; protein; 148 AA.
XX
AC   ADI35083;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human PLA2G1B protein sequence.
XX
KW   PLA2G1B; fat deposition; leanness; polymorphism;
KW   non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW   hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW   microalbuminuria; human.
XX
OS   Homo sapiens.
XX
PN   WO2004002295-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020830.
XX
PR   27-JUN-2002; 2002US-0392361P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML;
XX
DR   WPI; 2004-082843/08.
DR   N-PSDB; ADI35082.
XX
PT   Diagnosing a predisposition to fat deposition or leanness, useful for
PT   diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT   detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT   the subject.
XX
PS   Claim 1; SEQ ID NO 2; 91pp; English.
XX
CC   The invention relates to diagnosing a predisposition to fat deposition or
CC   leanness in a subject comprising detecting the presence or absence of a
CC   polymorphic variation associated with fat deposition at a polymorphic
CC   site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC   subject, where the presence of the polymorphic variation indicates a
CC   predisposition to fat deposition in the subject. The polymorphic
CC   variation is a guanine at position 7328 or thymine at position 9182 of
CC   the present sequence. The method is useful for diagnosing a
CC   predisposition to fat deposition or leanness in a subject, and
CC   consequently for diagnosing a predisposition to non-insulin dependent
CC   diabetes mellitus (NIDDM) in a subject and conditions such as
CC   hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC   hypercoagulability, or microalbuminuria, which can lead to early
CC   prescription of preventive measures. The present sequence represents a
CC   human PLA2G1B protein sequence.
XX
SQ   Sequence 148 AA;

Alignment Scores:
```

```
Pred. No.:              1e-22          Length:         148
Score:                  258.00         Matches:        55
Percent Similarity:     59.8%          Conservative:   6
Best Local Similarity:  53.9%          Mismatches:     12
Query Match:            20.6%          Indels:         29
DB:                     8              Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ADI35083 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
                ||||||||||   |||   |||:::                                |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                   :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                                        |||   :::||||||||
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                |:::
Db        108 rLys 109

RESULT 8
ADJ09984
ID   ADJ09984 standard; protein; 148 AA.
XX
AC   ADJ09984;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human phopholipase A2 (PLA2G1B) protein SeqID 2.
XX
KW   human; fat reduction; fat deposition; phospholipase A2; PLA2G1B;
KW   appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW   non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW   hypertension; antidiabetic.
XX
OS   Homo sapiens.
XX
PN   WO2004002296-A2.
XX
PD   08-JAN-2004.
XX
PF   27-JUN-2003; 2003WO-US020831.
XX
PR   27-JUN-2002; 2002US-0392362P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Adam GIR,  Langdown ML,  Denissenko MF,  Dennis E,  Cantor C;
PI   Rubin B;
XX
DR   WPI; 2004-071944/07.
DR   N-PSDB; ADJ09983.
XX
PT   Identifying a candidate therapeutic for fat reduction, useful for
PT   treating diabetes, by introducing a test molecule to a system comprising
PT   PLA2G1B protein or nucleic acid, and determining the presence of
PT   interaction between the compounds.
XX
PS   Claim 1; SEQ ID NO 2; 116pp; English.
XX
CC   This invention relates to a novel candidate therapeutic agent useful for
CC   fat reduction and disorders related to fat depositions. Specifically, it
CC   refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC   which is located on chromosome 12q24 and has been associated with central
CC   fat deposition. The present invention describes methods to detect the
CC   presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC   in particular G7328A and T9182G, and subsequently provide treatment that
CC   reduces fat deposition. This treatment may consist of an appetite
CC   suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC   regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC   psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC   deposition and it can be used to treat both obesity and non-insulin
CC   dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC   such as hypertension. As such, it exhibits antidiabetic activity. This
CC   polypeptide sequence is the human PLA2G1B protein of the invention.
XX
SQ   Sequence 148 AA;

Alignment Scores:
Pred. No.:              1e-22          Length:         148
Score:                  258.00         Matches:        55
Percent Similarity:     59.8%          Conservative:   6
Best Local Similarity:  53.9%          Mismatches:     12
Query Match:            20.6%          Indels:         29
DB:                     8              Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ADJ09984 (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
                ||||||||||   |||   |||:::                                |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                   :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                                        |||   :::||||||||
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                |:::
Db        108 rLys 109

RESULT 9
AAB54059
ID   AAB54059 standard; protein; 156 AA.
XX
AC   AAB54059;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen protein sequence SEQ ID NO:511.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
KW   diagnosis; identification; cytostatic; neuroprotective; nootropic;
KW   immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic; neural;
KW   immune system; muscular; reproductive; gastrointestinal; pulmonary;
KW   cardiovascular; renal; proliferative.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
```

```
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US005989.
XX
PR   12-MAR-1999;   99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   N-PSDB; AAC98824.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition.
XX
PS   Claim 11; Page 948; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention
XX
SQ   Sequence 156 AA;

Alignment Scores:
Pred. No.:              1.02e-22        Length:         156
Score:                  258.00          Matches:        55
Percent Similarity:     59.8%           Conservative:   6
Best Local Similarity:  53.9%           Mismatches:     12
Query Match:            20.6%           Indels:         29
DB:                     3               Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAB54059 (1-156)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA----------------------------------AGT 139
              ||||||||   |||   |||:::                                    |||
Db         27 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 46

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                ||| |||                 :::      :::||| |||   |||
Db         47 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 66

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
               |||:::                                    |||   :::||||||||
Db         66 rProValAsp----------------------------GluLeuAspLysCysCysGl 76

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 96

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 116

Qy        359 TAGG 362
                |:::
Db        116 rLys 117

RESULT 10
ABM83696
ID   ABM83696 standard; protein; 194 AA.
XX
AC   ABM83696;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human diagnostic and therapeutic pprotein SEQ ID NO:3945.
XX
KW   gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS   Homo sapiens.
XX
PN   WO2004023973-A2.
XX
PD   25-MAR-2004.
XX
PF   12-SEP-2003; 2003WO-US028227.
XX
PR   12-SEP-2002; 2002US-0410259P.
PR   12-SEP-2002; 2002US-0410260P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Schmidt JP,  Wright RJ,  Bruns CM,  Marjanovic MM,  Shen F;
PI   Harthshorne TA,  Suchorolski MT,  Altus CM,  Pitts SJ,  Elder LV;
PI   Mooney EM,  Delegeane AM,  Panesar IS,  Banville SC,  Reddy TP;
PI   Stevens KA,  Blanchard JL,  Panzer SR,  Wang X,  Au AP,  Gerstin EH;
PI   Peralta CH,  Anderson SB,  Rioux P,  Shen EJ,  Wu MC,  Stuve LL;
PI   Lagace RE,  Spiro PA,  Stewart EA,  Wingrove J,  Vitt UA,  Kirton ES;
PI   Xu Y,  Kwong M,  Policky JL,  Hurwitz BL,  Ma Y,  Jackson JL,  Gietzen D;
PI   Patury S,  Shi X,  Suarez CJ;
XX
DR   WPI; 2004-329368/30.
DR   N-PSDB; ACN42348.
XX
PT   New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT   in diagnosing a condition, disease or disorder associated with human
PT   molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT   in gene mapping.
XX
PS   Claim 27; Page; 190pp; English.
XX
CC   The invention relates to novel diagnostic and therapeutic polynucleotides
CC   selected from one of the 2722 sequences defined in the specification. A
CC   polynucleotide of the invention may have a use in gene therapy. The human
CC   diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC   used to diagnose a particular condition, disease or disorder associated
CC   with human molecules, e.g. cell proliferative disorders,
CC   autoimmune/inflammatory disorder, developmental disorder, endocrine
CC   disorder, neurological disorders, gastrointestinal disorders, or
CC   infections caused by virus, bacteria, fungi or parasite. The dithp
CC   molecules may also be used in genetic mapping, in identifying individuals
CC   from minute biological samples, in detecting single nucleotide
CC   polymorphisms, as molecular weight markers, and for somatic or germline
CC   gene therapy. The present sequence represents a dithp protein of the
CC   invention. Note: The sequence data for this patent is not represented in
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ   Sequence 194 AA;

Alignment Scores:
Pred. No.:              1.11e-22        Length:         194
Score:                  258.00          Matches:        55
Percent Similarity:     59.8%           Conservative:   6
```

```
Best Local Similarity:  53.9%          Mismatches:    12
Query Match:            20.6%          Indels:        29
DB:                     8              Gaps:          3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABM83696 (1-194)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         65 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 84

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                |||   |||                 :::     :::||| |||   |||
Db         85 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 104

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
               |||:::                                    |||   :::||||||||
Db        104 rProValAsp-----------------------------GluLeuAspLysCysCysGl 114

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 134

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 154

Qy        359 TAGG 362
              |:::
Db        154 rLys 155

RESULT 11
ABG11912
ID   ABG11912 standard; protein; 217 AA.
XX
AC   ABG11912;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11903.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS76099.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 42271; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 217 AA;

Alignment Scores:
Pred. No.:              1.16e-22        Length:         217
Score:                  258.00          Matches:        55
Percent Similarity:     59.8%           Conservative:   6
Best Local Similarity:  53.9%           Mismatches:     12
Query Match:            20.6%           Indels:         29
DB:                     4               Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABG11912 (1-217)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         88 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 107

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                |||   |||                 :::     :::||| |||   |||
Db        108 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 127

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
               |||:::                                    |||   :::||||||||
Db        127 rProValAsp-----------------------------GluLeuAspLysCysCysGl 137

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 157

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 177

Qy        359 TAGG 362
              |:::
Db        177 rLys 178

RESULT 12
ADO28599
ID   ADO28599 standard; protein; 145 AA.
XX
AC   ADO28599;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human PA21 protein SEQ ID NO:28.
XX
KW   high-grade dysplasia; HGD; oesophageal adenocarcinoma;
KW   neo-plastic transformation; cancer; cytostatic; gene therapy; human;
KW   phospholipase a2 precursor; PA21; chromosome 12.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 45
FT                   /note= "encoded by AACTAC"
FT   Misc-difference 101
```

```
FT                   /note= "encoded by GCAATC"
XX
PN   WO2004044178-A2.
XX
PD   27-MAY-2004.
XX
PF   13-NOV-2003; 2003WO-US036260.
XX
PR   13-NOV-2002; 2002US-0425813P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Smith V;
XX
DR   WPI; 2004-420319/39.
DR   N-PSDB; ADO28598.
XX
PT   Detecting of high-grade dysplasia in cells of a mammalian tissue sample
PT   comprises establishing the level of expression in the test tissue sample
PT   of the genes.
XX
PS   Example 4; SEQ ID NO 28; 256pp; English.
XX
CC   The present invention describes a method for detecting high-grade
CC   dysplasia (HGD) in cells of a mammalian tissue sample. Also described:
CC   (1) identifying an oesophageal tissue susceptible to oesophageal
CC   adenocarcinoma; (2) determining the predisposition of a mammalian tissue
CC   to a neo-plastic transformation by detecting HGD in cells of the tissue;
CC   and (3) detecting cancer in a patient. The method can be used in
CC   detecting HGD and cancer in cells of a mammalian tissue sample. The
CC   methods and compositions of the present invention can be used in treating
CC   and preventing HGD and cancer, and in gene therapy. The present sequence
CC   represents human phospholipase a2 precursor (PA21), which is used in the
CC   exemplification of the present invention. The human PA21 gene is located
CC   on chromosome 12.
XX
SQ   Sequence 145 AA;

Alignment Scores:
Pred. No.:              5.71e-21        Length:         145
Score:                  244.00          Matches:        54
Percent Similarity:     59.4%           Conservative:   6
Best Local Similarity:  53.5%           Mismatches:     12
Query Match:            19.5%           Indels:         29
DB:                     8               Gaps:           4

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ADO28599 (1-145)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         18 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 37

Qy        140 CATCCCCACCTG-----------------TTCCTGAGAGTGGGCAGGTGAGGGCTGACC 181
                 |||   |||                   :::      :::||| |||   |||   ||
Db         38 AspProPheLeuGluTyrAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyThrPr 57

Qy        182 TATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGAC 241
              |:::                                 |||   :::||||||||||
Db         57 oValAsp----------------------------GluLeuAspLysCysCysGlnTh 67

Qy        242 ACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAM 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 rHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAspAs 87

Qy        302 MCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAG 361
              |||||||||||||||||||||||||||||||||||   ||||||||||||||::
Db         87 nProTyrThrHisThrTyrSerTyrSerCysSerGlySer---IleThrCysSerSerLy 106

Qy        362 G 362
              :
Db        106 s 106

RESULT 13
AAB54421
ID   AAB54421 standard; protein; 79 AA.
XX
AC   AAB54421;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen protein sequence SEQ ID NO:873.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
KW   diagnosis; identification; cytostatic; neuroprotective; nootropic;
KW   immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic; neural;
KW   immune system; muscular; reproductive; gastrointestinal; pulmonary;
KW   cardiovascular; renal; proliferative.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US005989.
XX
PR   12-MAR-1999;   99US-0124270P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   N-PSDB; AAC99186.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition.
XX
PS   Claim 11; Page 1334; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention
XX
SQ   Sequence 79 AA;

Alignment Scores:
Pred. No.:              3.5e-19         Length:         79
Score:                  229.00          Matches:        39
Percent Similarity:     90.9%           Conservative:   1
Best Local Similarity:  88.6%           Mismatches:     4
Query Match:            18.3%           Indels:         0
DB:                     3               Gaps:           0
```

```
US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAB54421 (1-79)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   :::|||||||||||||||||||||||||||||||||||||||   |||||||||
Db         29 LeuAspLysCysCysGlnThrHisAspAsnCysTyrAspGlnAlaLys***LeuAspSer 48

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||   ||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db         49 Cys***PheLeuLeuAspAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySer 68

Qy        342 GCAATCACCTGT 353
              ||||||||||||
Db         69 AlaIleThrCys 72

RESULT 14
ABB08153
ID   ABB08153 standard; protein; 146 AA.
XX
AC   ABB08153;
XX
DT   10-SEP-2002  (first entry)
XX
DE   Murine PLA2 enzyme mGI polypeptide.
XX
KW   Phospholipase A2; GXII PLA2; phosphatidylethanolamine; Th2; GV PLA2;
KW   transgenic; immunosuppressive; antiallergic; cytostatic; antimicrobial;
KW   antidiabetic; antirheumatic; antiarthritic; antiinflammatory; mouse;
KW   neuroprotective; cerebroprotective; antiinfertility; contraceptive;
KW   mGXII-1 PLA2; enzyme; mGI.
XX
OS   Mus musculus.
XX
PN   WO200240655-A2.
XX
PD   23-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US044125.
XX
PR   06-NOV-2000; 2000US-0246316P.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Ho I,  Arm JP,  Austen KF,  Glimcher LH;
XX
DR   WPI; 2002-500219/53.
XX
PT   New group XII phospholipase A2 protein, useful for identifying modulators
PT   used for modulating prostaglandin production by Th2 cell and Th2 cell
PT   differentiation/activity, and treating allergy, cancer and type I
PT   diabetes.
XX
PS   Disclosure; Fig 1; 77pp; English.
XX
CC   The invention relates to an isolated group XII phospholipase A2 (GXII
CC   PLA2) protein or its biologically active portion, where the protein
CC   selectively hydrolyses arachidonic acid in sn-2 position of
CC   phosphatidylethanolamine. Methods of modulating Th2 cell differentiation
CC   activity by modulating either GXII PLA2 or GV PLA2 which is also
CC   preferentially expressed in T cells are provided. The GXII PLA2 proteins
CC   can be used to prepare anti-GXII PLA2 antibodies. The GXII PLA2 encoding
CC   nucleic acid molecules can be used to prepare non-human transgenic
CC   animals that contain cells carrying a transgene encoding GXII PLA2
CC   protein or a portion of GXII PLA2 protein. Prostaglandin production can
CC   be increased by stimulators of GXII PLA2 or GV PLA2 which is further
CC   useful for prolonging survival of the graft and thus has applications in
CC   bone marrow transplantation or solid organ transplantation, and for
CC   increasing production of Th2-promoting cytokines for commercial purposes.
CC   Modulating the type of T helper cell response mounted in the individual
CC   suffering from the disease condition e.g., inhibition of Th2 cell
CC   response using GXII PLA2 or GV PLA2 inhibitors is carried out for
CC   inhibiting prostaglandin production to thereby inhibit production of Th2-
CC   associated cytokines in: (a) allergic patients to downregulate production
CC   of pathogenic IgE antibodies; (b) cancer patients; and (c) subjects with
CC   infectious diseases. Stimulation of Th2 cell response using GXII PLA2 or
CC   GV PLA2 stimulators is useful for treating autoimmune diseases associated
CC   with Th2-type dysfunction, such as type I diabetes, rheumatoid arthritis,
CC   etc. The methods provided are also useful in treating conditions such as
CC   inflammation, arthritis, multiple sclerosis, stroke, infertility, and
CC   also have contraceptive uses. The present sequence represents a mouse
CC   PLA2 enzyme mGI
XX
SQ   Sequence 146 AA;

Alignment Scores:
Pred. No.:              3.32e-18        Length:         146
Score:                  222.00          Matches:        37
Percent Similarity:     91.5%           Conservative:   6
Best Local Similarity:  78.7%           Mismatches:     4
Query Match:            17.7%           Indels:         0
DB:                     5               Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x ABB08153 (1-146)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||||||||||:::||||||   ||||||||||||||:::|||
Db         63 LeuAspArgCysCysGlnThrHisAspHisCysTyrSerGlnAlaLysLysLeuGluSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||||||||||||:::|||   |||||||||:::|||||||||||||||||||||||||||
Db         83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrSerCysSerGlySer 102

Qy        342 GCAATCACCTGTAGCAGTAGG 362
              ||||||||||||::::::
Db        103 GluIleThrCysSerAlaLys 109

RESULT 15
AAR63059
ID   AAR63059 standard; protein; 124 AA.
XX
AC   AAR63059;
XX
DT   25-MAR-2003  (revised)
DT   15-AUG-1995  (first entry)
XX
DE   Rat PLA2 type I.
XX
KW   RPLA2; RPLA2-8; RPLA2-10; phospholipase A2; PLA2.
XX
OS   Homo sapiens.
XX
PN   WO9502328-A1.
XX
PD   26-JAN-1995.
XX
PF   15-JUL-1994;   94WO-US007926.
XX
PR   15-JUL-1993;   93US-00091941.
PR   26-JUL-1993;   93US-00097354.
XX
PA   (INDV ) UNIV INDIANA FOUND.
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tischfield JA,  Seilhamer JJ;
XX
DR   WPI; 1995-067096/09.
XX
PT   Novel type III and IV low mol. wt. phospholipase A2 enzymes - from humans
PT   and rats, also nucleic acid sequences useful, e.g. for recombinant prodn.
PT   of enzymes, research into Batten's disease, etc.
XX
PS   Disclosure; Fig 27; 160pp; English.
XX
CC   Rat cDNAs (AAQ81136-37) encoding 2 novel PLA2s, RPAL2-8 (AAR63044) and
CC   RPLA2-10 (AAR63045), were isolated from brain and heart libraries,
CC   respectively. RPAL2-8 and RPLA2-10 were characterized as novel type III
CC   and IV PLA2s on the basis of their cysteine content compared with RPLA2
```

```
CC   types I (AAR63059) and II (AAR63059). (Updated on 25-MAR-2003 to correct
CC   PN field.)
XX
SQ   Sequence 124 AA;

Alignment Scores:
Pred. No.:              1.77e-17        Length:         124
Score:                  216.00          Matches:        35
Percent Similarity:     89.4%           Conservative:   7
Best Local Similarity:  74.5%           Mismatches:     5
Query Match:            17.2%           Indels:         0
DB:                     2               Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x AAR63059 (1-124)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||||||||||||||:::||||||:::||||||||||||||:::|||
Db         41 LeuAspArgCysCysGlnThrHisAspHisCysTyrAsnGlnAlaLysLysLeuGluSer 60

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||||||||:::|||   |||||||||:::|||||||||||   |||||||||:::
Db         61 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrLysCysSerGlyAsn 80

Qy        342 GCAATCACCTGTAGCAGTAGG 362
                |||||||||||   :::
Db         81 ValIleThrCysSerAspLys 87


Search completed: February 8, 2006, 10:42:36
Job time : 103.179 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07 ; Search time 4.78382 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 1253
Sequence: 1 gtcgtgtcactgctgtccag.....:.....tgcagcctcaacctcctgag 701

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092020_27698/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_92_@runat_08022006_092020_27698 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	258	20.6	148	1	PSHU	phospholipase A2 (
	2	218	17.4	146	1	PSDG	phospholipase A2 (
	3	216	17.2	144	2	JN0480	phospholipase A2 (
	4	216	17.2	146	1	PSRT	phospholipase A2 (
	5	209	16.7	145	1	PSBOA	phospholipase A2 (
	6	205	16.4	146	1	PSPGA	phospholipase A2 (
	7	202	16.1	124	1	PSPGA2	phospholipase A2 (
	8	201.5	16.1	132	1	PSHOA	phospholipase A2 (
	9	188	15.0	146	1	S34049	phospholipase A2 (
	10	159	12.7	119	2	S00227	phospholipase A2 (
	11	157	12.5	125	2	S65624	phospholipase A(2)
	12	149.5	11.9	119	1	PSNJ3K	phospholipase A2 (
	13	144.5	11.5	146	1	PSNJAF	phospholipase A2 (
	14	143.5	11.5	119	1	PSNJ2K	phospholipase A2 (
	15	143	11.4	133	1	PSOXG	phospholipase A2 (
	16	140.5	11.2	119	1	PSRIA	phospholipase A2 (
	17	139.5	11.1	118	2	S01801	phospholipase A2 (
	18	139	11.1	117	2	A32416	phospholipase A2 (
	19	138.5	11.1	118	1	PSKFT3	phospholipase A2 (
	20	138.5	11.1	119	2	S07528	phospholipase A2 (
	21	137.5	11.0	118	1	PSKFT1	phospholipase A2 (
	22	137.5	11.0	118	1	PSKFT2	phospholipase A2 (
	23	136.5	10.9	119	2	JN0403	phospholipase A2 (
	24	136	10.9	118	2	G34860	phospholipase A2 (
	25	129.5	10.3	118	2	JK0223	phospholipase A2 (
	26	128.5	10.3	145	1	PSKF2U	phospholipase A2 (
	27	127	10.1	118	1	PSSNK1	phospholipase A2 (
	28	127	10.1	118	2	F34860	phospholipase A2 (
	29	126	10.1	118	1	PSNJ1M	phospholipase A2 (
	30	126	10.1	118	1	PSNJ2M	phospholipase A2 (
	31	125	10.0	119	1	PSNJ2W	phospholipase A2 (
	32	125	10.0	119	1	PSNJ3W	phospholipase A2 (
	33	123.5	9.9	118	1	PSNJ1W	phospholipase A2 (
	34	122	9.7	118	2	A32622	phospholipase A2 (
	35	119	9.5	118	1	PSNJ3B	phospholipase A2 (
	36	118	9.4	118	1	PSNJ3M	phospholipase A2 (
	37	116	9.3	117	2	A34860	phospholipase A2 (
	38	114	9.1	121	2	S29652	textilotoxin chain
	39	112.5	9.0	119	1	PSNOA1	phospholipase A2 h
	40	112.5	9.0	119	1	PSOXA	phospholipase A2 (
	41	112	8.9	137	2	JC4710	beta-bungarotoxin
	42	111	8.9	118	2	E34860	phospholipase A2 (
c	43	110.5	8.8	613	4	C40201	artifact-warning s
	44	110	8.8	145	1	PSKFA2	phospholipase A2 (
	45	108	8.6	613	4	C40201	artifact-warning s

ALIGNMENTS

RESULT 1
PSHU
phospholipase A2 (EC 3.1.1.4) IB precursor [validated] - human
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Homo sapiens (man)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: C25793; A25793; A00733; A21566
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986
A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcin
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: C25793
A;Molecule type: DNA
A;Residues: 1-148 <SEI2>
A;Cross-references: UNIPROT:P04054; UNIPARC:UPI0000048FD9; GB:M22970
A;Accession: A25793
A;Molecule type: mRNA
A;Residues: 1-148 <SEI>
A;Cross-references: UNIPARC:UPI0000048FD9; GB:M21055
A;Experimental source: lung
R;Grataroli, R.; Dijkman, R.; Dutilh, C.E.; Van der Ouderaa, F.; De Haas, G.H.; Figarel
Eur. J. Biochem. 122, 111-117, 1982
A;Title: Studies on prophospholipase A2 and its enzyme from human pancreatic juice. Cat
A;Reference number: A91114; MUID:82138816; PMID:7060561
A;Accession: A00733
A;Molecule type: protein
A;Residues: 16-22 <VER1>
A;Cross-references: UNIPARC:UPI000014C3BF
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A21566
A;Molecule type: protein
A;Residues: 23-143,'YS',146-147 <VER2>
A;Cross-references: UNIPARC:UPI0000172758
C;Genetics:
A;Gene: GDB:PLA2G1B; PLA2A; PLA2; P

A;Cross-references: GDB:120715; OMIM:172410
A;Map position: 12q23-12qter
A;Introns: 12/1; 65/2; 108/1
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; met
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <ACT>
F;23-148/Product: phospholipase A2 IB #status experimental <MAT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.:	2.99e-20	Length:	148
Score:	258.00	Matches:	55
Percent Similarity:	59.8%	Conservative:	6
Best Local Similarity:	53.9%	Mismatches:	12
Query Match:	20.6%	Indels:	29
DB:	1	Gaps:	3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSHU (1-148)

```
Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||   |||                        :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                                    |||   :::|||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109
```

RESULT 2
PSDG
phospholipase A2 (EC 3.1.1.4) precursor - dog
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Canis lupus familiaris (dog)
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Aug-2004
C;Accession: S11316; A24392; JS0006
R;Kerfelec, B.; LaForge, K.S.; Vasiloudes, P.; Puigserver, A.; Scheele, G.A.
Eur. J. Biochem. 190, 299-304, 1990
A;Title: Isolation and sequence of the canine pancreatic phospholipase A(2) gene.
A;Reference number: S11316; MUID:90306027; PMID:2142076
A;Accession: S11316
A;Molecule type: DNA
A;Residues: 1-146 <KE2>
A;Cross-references: UNIPROT:P06596; UNIPARC:UPI000013108C
R;Kerfelec, B.; LaForge, K.S.; Puigserver, A.; Scheele, G.
Pancreas 1, 430-437, 1986
A;Title: Primary structures of canine pancreatic lipase and phospholipase A2 messenger R
A;Reference number: A93751; MUID:87175472; PMID:3562437
A;Accession: A24392
A;Molecule type: mRNA
A;Residues: 1-146 <KER>
A;Cross-references: UNIPARC:UPI000013108C; GB:M35301; NID:g164041; PIDN:AAA30883.1; PID:
R;Ohara, O.; Tamaki, M.; Nakamura, E.; Tsuruta, Y.; Fujii, Y.; Shin, M.; Teraoka, H.; Ok

J. Biochem. 99, 733-739, 1986
A;Title: Dog and rat pancreatic phospholipases A2: complete amino acid sequences deduced
A;Reference number: A92008; MUID:86223862; PMID:3754861
A;Accession: JS0006
A;Molecule type: mRNA
A;Residues: 1-146 <OHA>
A;Cross-references: UNIPARC:UPI000013108C; GB:D00035; GB:N00035; NID:g217659; PIDN:BAA00(
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-s
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micel
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; meta
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <APT>
F;23-146/Product: phospholipase A2 #status predicted <MPT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.:	8.9e-16	Length:	146
Score:	218.00	Matches:	48
Percent Similarity:	54.9%	Conservative:	8
Best Local Similarity:	47.1%	Mismatches:	17
Query Match:	17.4%	Indels:	29
DB:	1	Gaps:	3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSDG (1-146)

```
Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         19 IleSerProArgAlaValTrpGlnPheArgAsnMetIleLysCysThrIleProGluSer 38

Qy        140 CATCCC--------------------CACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  |||                        :::      :::||| |||   |||
Db         39 AspProLeuLysAspTyrAsnAspTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                                    |||   :::|||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||:::||||||   :::||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspHisCysTyrSerGluAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||       ||||||||||||||||||||||   ||||||||||||||
Db         88 pAsnProTyrThrLysIleTyrSerTyrSerCysSerGlySerGluIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109
```

RESULT 3
JN0480
phospholipase A2 (EC 3.1.1.4) precursor, pancreatic type - rabbit
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Oct-2004
C;Accession: JN0480
R;Kumar, V.B.
Biochem. Biophys. Res. Commun. 192, 683-692, 1993
A;Title: Cloning and expression of rabbit pancreatic phospholipase A2.
A;Reference number: JN0480; MUID:93249443; PMID:7916604
A;Accession: JN0480
A;Molecule type: DNA
A;Residues: 1-144 <KUM>
A;Cross-references: UNIPROT:Q7M334; UNIPARC:UPI000017586E; GB:L11899
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid digestion; pancreas; zymogen
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <PRO>
F;23-144/Product: phospholipase A2, pancreatic type #status predicted <MAT>

```
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:              1.49e-15      Length:          144
Score:                  216.00        Matches:         45
Percent Similarity:     55.9%         Conservative:    12
Best Local Similarity:  44.1%         Mismatches:      16
Query Match:            17.2%         Indels:          29
DB:                     2             Gaps:            3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x JN0480 (1-144)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              :::||||||   |||   |||:::                                 |||
Db         19 ValSerProThrAlaLeuTrpGlnPheArgGlyMetIleGlnCysThrIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                 |||:::|||                    :::      :::||| |||   |||
Db         39 SerProTyrLeuGluPheAsnGlyTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                |||:::                          |||   ||||||||||.
Db         58 rProValAsp----------------------------GluLeuAspArgCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||   ||||||   ||||||||||||||   ||||||  ||||||:::||
Db         68 nThrHisAspGlnCysTyrThrGlnAlaLysLysLeuSerSerCysSerPheLeuValAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||:::::|||||||||||||||||||||:::   :::||||||||||
Db         88 pAsnProTyrThrAsnSerTyrSerTyrSerCysSerGlyThrThrValThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109

RESULT 4
PSRT
phospholipase A2 (EC 3.1.1.4) precursor - rat
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Aug-2004
C;Accession: A92008; S02147; A91987; A28617; A05322; JS0007
R;Ohara, O.; Tamaki, M.; Nakamura, E.; Tsuruta, Y.; Fujii, Y.; Shin, M.; Teraoka, H.; Ok
J. Biochem. 99, 733-739, 1986
A;Title: Dog and rat pancreatic phospholipases A2: complete amino acid sequences deduced
A;Reference number: A92008; MUID:86223862; PMID:3754861
A;Accession: A92008
A;Molecule type: mRNA
A;Residues: 1-146 <OHA>
A;Cross-references: UNIPROT:P04055; UNIPARC:UPI000013109F; GB:D00036; GB:N00036; NID:g22
R;Sakata, T.; Nakamura, E.; Tsuruta, Y.; Tamaki, M.; Teraoka, H.; Tojo, H.; Ono, T.; Oka
Biochim. Biophys. Acta 1007, 124-126, 1989
A;Title: Presence of pancreatic-type phospholipase A(2) mRNA in rat gastric mucosa and l
A;Reference number: S02147; MUID:89076874; PMID:2909239
A;Accession: S02147
A;Molecule type: mRNA
A;Residues: 1-146 <SAK>
A;Cross-references: UNIPARC:UPI000013109F; EMBL:D00036; NID:g220874; PIDN:BAA00024.1; PI
R;Ono, T.; Tojo, H.; Inoue, K.; Kagamiyama, H.; Yamano, T.; Okamoto, M.
J. Biochem. 96, 785-792, 1984
A;Title: Rat pancreatic phospholipase A2: purification, characterization, and N-terminal
A;Reference number: A91987; MUID:85054750; PMID:6501264
A;Accession: A91987
A;Molecule type: protein
A;Residues: 23-54 <ONO>
A;Cross-references: UNIPARC:UPI0000172759
R;Tojo, H.; Ono, T.; Kuramitsu, S.; Kagamiyama, H.; Okamoto, M.
J. Biol. Chem. 263, 5724-5731, 1988
A;Title: A phospholipase A-2 in the supernatant fraction of rat spleen. Its similarity t
A;Reference number: A28617; MUID:88186889; PMID:3128547
A;Accession: A28617
A;Molecule type: protein
A;Residues: 23-54 <TOJ>
A;Cross-references: UNIPARC:UPI0000172759
A;Experimental source: spleen
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; me
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <APT>
F;23-146/Product: phospholipase A2 #status experimental <MAT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:              1.49e-15      Length:          146
Score:                  216.00        Matches:         35
Percent Similarity:     89.4%         Conservative:    7
Best Local Similarity:  74.5%         Mismatches:      5
Query Match:            17.2%         Indels:          0
DB:                     1             Gaps:            0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSRT (1-146)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||||||||||||:::||||||:::|||||||||||||||:::|||
Db         63 LeuAspArgCysCysGlnThrHisAspHisCysTyrAsnGlnAlaLysLysLeuGluSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||||||||||||:::|||   |||||||||:::|||||||||||   |||||||||:::
Db         83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrLysCysSerGlyAsn 102

Qy        342 GCAATCACCTGTAGCAGTAGG 362
                  ||||||||||||   :::
Db        103 ValIleThrCysSerAspLys 109

RESULT 5
PSBOA
phospholipase A2 (EC 3.1.1.4) precursor [validated] - bovine
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Nov-1980 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: A27508; A94661; A91252; A00736
R;Tanaka, T.; Kimura, S.; Ota, Y.
Nucleic Acids Res. 15, 3178, 1987
A;Title: Sequence of a cDNA coding for bovine pancreatic phospholipase A2.
A;Reference number: A27508; MUID:87174835; PMID:3562249
A;Accession: A27508
A;Molecule type: mRNA
A;Residues: 1-145 <TAN>
A;Cross-references: UNIPROT:P00593; UNIPARC:UPI0000167C39; EMBL:Y00120; NID:g634; PIDN:C
R;Dutilh, C.E.; Van Doren, P.J.; Verheul, F.E.A.M.; De Haas, G.H.
Eur. J. Biochem. 53, 91-97, 1975
A;Title: Isolation and properties of prophospholipase A2 from ox and sheep pancreas.
A;Reference number: A94661
A;Accession: A94661
A;Molecule type: protein
A;Residues: 16-22 <DUT>
A;Cross-references: UNIPARC:UPI000017275A
R;Fleer, E.A.M.; Verheij, H.M.; de Haas, G.H.
Eur. J. Biochem. 82, 261-269, 1978
A;Title: The primary structure of bovine pancreatic phospholipase A2.
A;Reference number: A91252; MUID:78084411; PMID:620674
A;Accession: A91252
A;Molecule type: protein
A;Residues: 23-143,'N',145 <FLE>
A;Cross-references: UNIPARC:UPI00001109B6
R;Dijkstra, B.W.; Drenth, J.; Kalk, K.H.; Vandermaelen, P.J.
J. Mol. Biol. 124, 53-60, 1978
A;Title: Three-dimensional structure and disulfide bond connections in bovine pancreati
A;Reference number: A92853; MUID:79050564; PMID:712836
A;Contents: annotation; X-ray crystallography, 2.4 angstroms; disulfide bonds
```

R;Dijkstra, B.W.; Drenth, J.; Kalk, K.H.
Nature 289, 604-606, 1981
A;Title: Active site and catalytic mechanism of phospholipase A2.
A;Reference number: A93243; MUID:81123082; PMID:7464926
A;Contents: annotation; X-ray crystallography, 1.7 angstroms; active site; catalytic mec
R;Dijkstra, B.W.; van Nes, G.J.H.; Kalk, K.H.; Brandenburg, N.P.; Hol, W.G.J.; Drenth, J
submitted to the Brookhaven Protein Data Bank, June 1981
A;Reference number: A50406; PDB:2BP2
A;Contents: annotation; X-ray crystallography, 3.0 angstroms, residues 23-143,'N',145
A;Note: pancreas, phospholipase A2 precursor
R;Dijkstra, B.W.; van Nes, G.J.H.; Kalk, K.H.; Brandenburg, N.P.; Hol, W.G.J.; Drenth, J
Acta Crystallogr. B38, 793-799, 1982
A;Title: The structure of bovine pancreatic prophospholipase A2 at 3.0 angstroms resolut
A;Reference number: A58516
A;Contents: annotation; X-ray crystallography, 3.0 angstroms
R;Sundaralingam, M.
submitted to the Brookhaven Protein Data Bank, January 1992
A;Reference number: A51455; PDB:2BPP
A;Contents: annotation; X-ray crystallography, 1.8 angstroms, residues 23-143,'N',145
A;Note: recombinant form expressed in Escherichia coli
R;Noel, J.P.; Bingman, C.A.; Deng, T.; Dupureur, C.M.; Hamilton, K.J.; Jiang, R.T.; Kwak
Biochemistry 30, 11801-11811, 1991
A;Title: Phospholipase A2 engineering. X-ray structural and functional evidence for the
A;Reference number: A58513; MUID:92089090; PMID:1751497
A;Contents: annotation; X-ray crystallography, 1.8 angstroms
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; met
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-145/Product: prophospholipase A2 #status experimental <PRO>
F;16-22/Domain: activation peptide #status experimental <APT>
F;23-145/Product: phospholipase A2 #status experimental <MAT>
F;16/Modified site: pyrrolidone carboxylic acid (Gln) (in precursor form) #link PRO #sta
F;33-99,49-145,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status experimental
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status experimental
F;70,121/Active site: His, Asp #status experimental

Alignment Scores:

Pred. No.:	9.04e-15	Length:	145
Score:	209.00	Matches:	36
Percent Similarity:	84.8%	Conservative:	3
Best Local Similarity:	78.3%	Mismatches:	7
Query Match:	16.7%	Indels:	0
DB:	1	Gaps:	0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSBOA (1-145)

```
Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         63 LeuAspArgCysCysGlnThrHisAspAsnCysTyrLysGlnAlaLysLysLeuAspSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||   |||:::|||   |||||||||:::   ||||||||||||||||||   :::
Db         83 CysLysValLeuValAspAsnProTyrThrAsnAsnTyrSerTyrSerCysSerAsnAsn 102

Qy        342 GCAATCACCTGTAGCAGT 359
              ||||||||||||||||
Db        103 GluIleThrCysSerSer 108
```

RESULT 6
PSPGA
phospholipase A2 (EC 3.1.1.4) major precursor [validated] - pig
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1993 #text_change 16-Aug-2004
C;Accession: B25793; S30750; A60264; A00734
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986
A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: B25793
A;Molecule type: mRNA
A;Residues: 1-146 <SEI>
A;Cross-references: UNIPROT:P00592; UNIPARC:UPI000013109C; EMBL:M21055; NID:g164611; PIDN
A;Experimental source: pancreas
R;de Geus, P.; van den Bergh, C.J.; Kuipers, O.; Verheij, H.M.; Hoekstra, W.P.M.; de Haas
Nucleic Acids Res. 15, 3743-3759, 1987
A;Title: Expression of porcine pancreatic phospholipase A(2). Generation of active enzyme
A;Reference number: S30750; MUID:87231066; PMID:3295782
A;Accession: S30750
A;Molecule type: mRNA
A;Residues: 1-146 <DEG>
A;Cross-references: UNIPARC:UPI000013109C; EMBL:Y00146; NID:g2051; PIDN:CAA68341.1; PID:g
R;Kuipers, O.P.; van den Bergh, C.J.; Verheij, H.M.; de Haas, G.H.
Adv. Exp. Med. Biol. 279, 65-84, 1990
A;Title: Probing the mechanism of pancreatic phospholipase A2 with the aid of recombinant
A;Reference number: A60264; MUID:91263889; PMID:2096701
A;Accession: A60264
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-146 <KUI>
A;Cross-references: UNIPARC:UPI000013109C
R;Puijk, W.C.; Verheij, H.M.; de Haas, G.H.
Biochim. Biophys. Acta 492, 254-259, 1977
A;Title: The primary structure of phospholipase A-2 from porcine pancreas. A reinvestigat
A;Reference number: A90620; MUID:77222092; PMID:884127
A;Contents: revision
A;Accession: A00734
A;Molecule type: protein
A;Residues: 16-146 <PUI>
A;Cross-references: UNIPARC:UPI0000111BE3
R;de Haas, G.H.; Slotboom, A.J.; Bonsen, P.P.M.; Nieuwenhuizen, W.; van Deenen, L.L.M.; M
Biochim. Biophys. Acta 221, 54-61, 1970
A;Title: Studies on phospholipase A and its zymogen from porcine pancreas. II. The assign
A;Reference number: A90579; MUID:71014236; PMID:4919729
A;Contents: annotation; disulfide bonds
R;Dijkstra, B.W.; Renetseder, R.; Kalk, K.H.; Hol, W.G.J.; Drenth, J.
submitted to the Brookhaven Protein Data Bank, June 1983
A;Reference number: A50305; PDB:1P2P
A;Contents: annotation; X-ray crystallography, 2.6 angstroms
R;Dijkstra, B.W.; Renetseder, R.; Kalk, K.H.; Hol, W.G.J.; Drenth, J.
J. Mol. Biol. 168, 163-179, 1983
A;Title: Structure of porcine pancreatic phospholipase A-2 at 2.6 angstrom resolution and
A;Reference number: A92899; MUID:83268704; PMID:6876174
A;Contents: annotation; X-ray crystallography, 2.6 angstroms; active and binding sites
R;van den Bergh, C.J.; Slotboom, A.J.; Verheij, H.M.; de Haas, G.H.
J. Cell. Biochem. 39, 379-390, 1989
A;Title: The role of Asp-49 and other conserved amino acids in phospholipases A2 and the
A;Reference number: A60719; MUID:89255682; PMID:2722967
A;Contents: annotation
R;Finzel, B.C.; Ohlendorf, D.H.; Weber, P.C.; Salemme, F.R.
submitted to the Brookhaven Protein Data Bank, October 1991
A;Reference number: A50680; PDB:4P2P
A;Contents: annotation; X-ray crystallography, 2.4 angstroms, residues 23-146
R;Finzel, B.C.; Ohlendorf, D.H.; Weber, P.C.; Salemme, F.R.
Acta Crystallogr. B 47, 558-559, 1991
A;Title: An independent crystallographic refinement of porcine phospholipase A2 at 2.4 a
A;Reference number: A58515; MUID:92029796; PMID:1930837
A;Contents: annotation; X-ray crystallography, 2.4 angstroms
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; met
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <APT>
F;23-146/Product: phospholipase A2 major #status experimental <MAT>
F;16/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status experimental
F;50,52,54,71/Binding site: calcium, high affinity (Tyr, Gly, Gly, Asp) #status experime
F;70,121/Active site: His, Asp #status experimental
F;93,94,114/Binding site: calcium, low affinity (Glu, Ser, Glu) #status predicted

Alignment Scores:

```
Pred. No.:              2.53e-14        Length:         146
Score:                  205.00          Matches:        44
Percent Similarity:     53.9%           Conservative:   11
Best Local Similarity:  43.1%           Mismatches:     18
Query Match:            16.4%           Indels:         29
DB:                     1               Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSPGA (1-146)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA------------------------------AGT 139
              |||||||      |||   |||:::                               |||
Db         19 IleSerSerArgAlaLeuTrpGlnPheArgSerMetIleLysCysAlaIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
              ||||||   :::                    :::      :::||| |||   |||
Db         39 HisProLeuMetAspPheAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
              |||:::                                 |||   |||||||||::
Db         58 rProValAsp----------------------------GluLeuAspArgCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              :|||||||||||||||||      ||||||   ||||||||||||||||||:::||
Db         68 uThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSerCysLysPheLeuValAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||   :::||||||||||||||||||   :::   |||||||||:::||
Db         88 pAsnProTyrThrGluSerTyrSerTyrSerCysSerAsnThrGluIleThrCysAsnSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109

RESULT 7
PSPGA2
phospholipase A2 (EC 3.1.1.4) minor - pig (tentative sequence)
N;Alternate names: pancreatic phospholipase A2 minor isozyme; phosphatidylcholine 2-acyl
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Aug-2004
C;Accession: A00735
R;Puijk, W.C.; Verheij, H.M.; Wietzes, P.; de Haas, G.H.
Biochim. Biophys. Acta 580, 411-415, 1979
A;Title: The amino acid sequence of the phospholipase A-2 isoenzyme from porcine pancrea
A;Reference number: A00735; MUID:80088382; PMID:518908
A;Accession: A00735
A;Molecule type: protein
A;Residues: 1-124 <PUI>
A;Cross-references: UNIPROT:P04416; UNIPARC:UPI00001310BD
A;Experimental source: pancreas
C;Comment: This isozyme constitutes about 5% of the pancreatic phospholipase.
C;Comment: Phospholipase A2 occurs in virtually every mammalian tissue that has been exa
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; met
F;11-77,27-124,29-45,44-105,51-98,61-91,84-96/Disulfide bonds: #status experimental
F;28,30,32,49/Binding site: calcium (Tyr, Gly, Gly, Asp) #status experimental
F;48,99/Active site: His, Asp #status experimental

Alignment Scores:
Pred. No.:              5.55e-14        Length:         124
Score:                  202.00          Matches:        33
Percent Similarity:     85.1%           Conservative:   7
Best Local Similarity:  70.2%           Mismatches:     7
Query Match:            16.1%           Indels:         0
DB:                     1               Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSPGA2 (1-124)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||:::|||||||||||||||||||      ||||||   |||||||||
Db         41 LeuAspArgCysCysGluThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSer 60

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||||||||||:::|||   |||||||||::::::||||||||||||||||||   :::
Db         61 CysLysPheLeuValAspAsnProTyrThrAsnSerTyrSerTyrSerCysSerAsnThr 80

Qy        342 GCAATCACCTGTAGCAGTAGG 362
              |||||||||:::|||:::
Db         81 GluIleThrCysAsnSerLys 87

RESULT 8
PSHOA
phospholipase A2 (EC 3.1.1.4) precursor - horse (tentative sequence)
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Equus caballus (domestic horse)
C;Date: 24-Apr-1984 #sequence_revision 03-Aug-1984 #text_change 16-Aug-2004
C;Accession: A90615; A92198; A90651; A00737
R;Evenberg, A.; Meyer, H.; Verheij, H.M.; de Haas, G.H.
Biochim. Biophys. Acta 491, 265-274, 1977
A;Title: Isolation and properties of prophospholipase A-2 and phospholipase A-2 from ho
A;Reference number: A90615; MUID:77134902; PMID:849461
A;Accession: A90615
A;Molecule type: protein
A;Residues: 1-7 <EVE>
A;Cross-references: UNIPROT:P00594; UNIPARC:UPI000017275B
A;Note: a second form, lacking residues 1 and 2, was also found
R;Evenberg, A.; Meyer, H.; Gaastra, W.; Verheij, H.M.; de Haas, G.H.
J. Biol. Chem. 252, 1189-1196, 1977
A;Title: Amino acid sequence of phospholipase A-2 from horse pancreas.
A;Reference number: A92198; MUID:77118587; PMID:838712
A;Contents: active site
A;Accession: A92198
A;Molecule type: protein
A;Residues: 8-131 <EV2>
A;Cross-references: UNIPARC:UPI000017275C
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A90651
A;Molecule type: protein
A;Residues: 1-132 <VER>
A;Cross-references: UNIPARC:UPI0000131091
C;Comment: Phospholipase A2 occurs in virtually every mammalian tissue that has been ex
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; me
F;1-7/Domain: activation peptide #status experimental <APT>
F;8-132/Product: phospholipase A2 #status experimental <MPT>
F;18-84,34-131,36-52,51-112,58-105,68-98,91-103/Disulfide bonds: #status predicted
F;35,37,39,56/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;55,106/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:              6.28e-14        Length:         132
Score:                  201.50          Matches:        48
Percent Similarity:     43.5%           Conservative:   19
Best Local Similarity:  31.2%           Mismatches:     31
Query Match:            16.1%           Indels:         56
DB:                     1               Gaps:           6

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSHOA (1-132)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA------------------------------AGT 139
              |||||||||   |||   |||:::                               |||
Db          4 IleSerProArgAlaValTrpGlnPheArgSerMetIleGlnCysThrIleProAsnSer 23

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
              |||:::|||                      :::      :::||| |||   |||
Db         24 LysProTyrLeuGluPheAsnAspTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 43

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
```

```
              |||:::                                    |||      ||||||||
Db         43 rProValAsp-----------------------------GluLeuAspAlaCysCysGl 53

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              |   |||||||||||||||   |||||||||:::|||   ||||||:::||||||:::||
Db         53 nValHisAspAsnCysTyrThrGlnAlaLysGluLeuSerSerCysArgPheLeuValAs 73

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||   :::|||   :::|||||||||||||:::   :::|||||||||
Db         73 pAsnProTyrThrGluSerTyrLysPheSerCysSerGlyThrGluValThrCysSerAs 93

Qy        359 TAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAA 418
              :::                                                       ||
Db         93 pLys-------------------------------------------------------As 95

Qy        419 TAATAGT-----------------AACAACAGCCATGATTTAGTGTTAATTTTCTTGGT 460
              ||||:::                 |||         :::      :::   |||
Db         95 nAsnAlaCysGluAlaPheIleCysAsnCysAspArgAsnAlaAlaIleCysPhe----- 113

Qy        461 TCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 500
                        :::||||||:::|||      :::
Db        114 ----SerLysAlaProTyrAsnProGluAsnLysAsnLeu 125

RESULT 9
S34049
phospholipase A2 (EC 3.1.1.4) precursor, gastric and pancreatic [validated] - guinea pig
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Cavia porcellus (guinea pig)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 16-Aug-2004
C;Accession: S34049; S34046; S34048; S34047
R;Ying, Z.; Tojo, H.; Nonaka, Y.; Okamoto, M.
Eur. J. Biochem. 215, 91-97, 1993
A;Title: Cloning and expression of phospholipase A(2) from guinea pig gastric mucosa, it
A;Reference number: S34049; MUID:93345504; PMID:8344290
A;Accession: S34049
A;Molecule type: mRNA
A;Residues: 1-146 <YIN>
A;Cross-references: UNIPROT:P43434; UNIPARC:UPI000013108D; GB:D00740; NID:g398128; PIDN:
R;Tojo, H.; Ying, Z.; Okamoto, M.
Eur. J. Biochem. 215, 81-90, 1993
A;Title: Purification and characterization of guinea pig gastric phospholipase A(2) of t
A;Reference number: S34046; MUID:93345502; PMID:8344288
A;Accession: S34046
A;Status: preliminary
A;Molecule type: protein
A;Residues: 23-63 <TOJ>
A;Cross-references: UNIPARC:UPI00000E7305
A;Experimental source: gastric mucosa
A;Accession: S34048
A;Status: preliminary
A;Molecule type: protein
A;Residues: 23-24,'X',26-32,'X',34 <TO2>
A;Cross-references: UNIPARC:UPI000017275D
A;Experimental source: pancreas
C;Comment: In this species phospholipase activity is high in gastric mucosa and jueic an
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; gastric juice; lipid degradation; lipid
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status predicted <ACT>
F;23-146/Product: phospholipase A2 #status experimental <MAT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Phe, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:              2.02e-12        Length:         146
Score:                  188.00          Matches:        41
Percent Similarity:     53.2%           Conservative:   17
Best Local Similarity:  37.6%           Mismatches:     22
Query Match:            15.0%           Indels:         29
DB:                     1               Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S34049 (1-146)

Qy         92 ATTTTTTCAAGCCATAGGATCATTTCTCCTGAAGCATCTTGGCGA---------------- 136
              :::   ::::::      ||||||:::|||   |||   |||:::
Db         12 ValSerAlaAlaAlaHisIleIleThrProArgAlaLeuTrpGlnPheArgAspMetIle 31

Qy        137 ------------------AGTCATCCCCAC----------------------CTGTTCCTG 157
                                 |||   |||:::                      |||
Db         32 LysCysAlaIleProGlySerArgProTyrSerGluTyrAsnAsnTyrGlyCysPheCys 51

Qy        158 AGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCT 217
                 :::||| |||   |||   |||:::
Db         52 GlyLeuGly-GlySerGlyThrProValAsp---------------------------- 61

Qy        218 CCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGA 277
              |||   |||||||||:::   ||||||   ||||||   |||||||||   |||::
Db         62 -GluLeuAspArgCysCysGluIleHisAspAlaCysTyrThrGlnAlaLysHisLeuGl 81

Qy        278 CAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGG 337
              :|||||||||   ::::::|||   |||||||||:::::|||||:::|||||||||||
Db         81 uSerCysLysSerValIleAspAsnProTyrThrAsnSerTyrSerPheSerCysSerGl 101

Qy        338 CTCGGCAATCACCTGTAGCAGTAGG 362
              |:::   |||   ||||||||:::
Db        101 yThrAsnIleIleCysSerSerLys 109

RESULT 10
S00227
phospholipase A2 (EC 3.1.1.4) - shield snake
C;Species: Aspidelaps scutatus (shield snake)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 05-Oct-2004
C;Accession: S00227
R;Joubert, F.J.
Biol. Chem. Hoppe-Seyler 368, 1597-1602, 1987
A;Title: Purification, some properties of two phospholipases A2 (CM-I and CM-II) and the
A;Reference number: S00227; MUID:88163080; PMID:3442602
A;Accession: S00227
A;Molecule type: protein
A;Residues: 1-119 <JOU>
A;Cross-references: UNIPROT:P07037; UNIPARC:UPI00001310AB
A;Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 61-Leu
C;Comment: Phospholipase A2 hydrolyzes the 2-acyl groups in 3-sn-phosphoglycerides.
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; homodimer; lipid degradation; venom
F;1-119/Product: phospholipase A2 #status experimental <MAT>
F;11-71,26-118,28-44,43-99,50-92,60-85,78-90/Disulfide bonds: #status predicted
F;47,93/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:              3.58e-09        Length:         119
Score:                  159.00          Matches:        38
Percent Similarity:     41.1%           Conservative:   6
Best Local Similarity:  35.5%           Mismatches:     25
Query Match:            12.7%           Indels:         38
DB:                     2               Gaps:           5

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S00227 (1-119)

Qy         66 AATAAATATATGATAAAAGGACTTATA----------------------------TTT 95
              |||   |||       |||   :::|||
Db          1 AsnLeuTyrGlnPheLysAsnMetIleGlnCysThrValProAsnArgSerTrpTrpHis 20

Qy         96 TTTCAAGCCATAGGATCATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCC 155
              |||         |||   |||                                   |||
Db         21 PheAlaAspTyrGlyCysPhe--------------------------------CysGly 29

Qy        156 TGAGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCC 215
                    :::   |||   |||   |||:::
Db         30 ------TyrGlyGlySerGlyThrProValAsp------------------------- 38

Qy        216 CTCCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTG 275
```

```
                    |||   |||||||||||||||||||||||||||||||   :::|||:::||||||
Db          39 ---GluLeuAspArgCysCysGlnThrHisAspAsnCysTyrSerGluAlaGluLysLeu 57

Qy         276 GACAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCT 335
                |||||||                |||||||      ||||||||||||   ||||||
Db          58 SerGlyCysLys--------------ProTyrIleLysThrTyrSerTyrAspCysSer 72

Qy         336 GGCTCGGCAATCACCTGTAGC 356
                  :::||||||||||
Db          73 GlnGlyLysLeuThrCysSer 79
```

RESULT 11
S65624
phospholipase A(2) - common tiger snake
C;Species: Notechis scutatus scutatus (common tiger snake, mainland tiger snake)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Oct-2004
C;Accession: S65624
R;Francis, B.; Coffield, J.A.; Simpson, L.L.; Kaiser, I.I.
Arch. Biochem. Biophys. 318, 481-488, 1995
A;Title: Amino acid sequence of a new type of toxic phospholipase A(2) from the venom of
A;Reference number: S65624; MUID:95251401; PMID:7733680
A;Accession: S65624
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-125 <FRA>
A;Cross-references: UNIPROT:Q9PSN5; UNIPARC:UPI000013113B
C;Superfamily: Phospholipase A2
F;48,99/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.:	5.97e-09	Length:	125
Score:	157.00	Matches:	23
Percent Similarity:	75.0%	Conservative:	10
Best Local Similarity:	52.3%	Mismatches:	11
Query Match:	12.5%	Indels:	0
DB:	2	Gaps:	0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x S65624 (1-125)

```
Qy         222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
               |||   |||||||||||||||||||||:::||||||   :::|||:::||||||   :::
Db          41 LeuAspArgCysCysGlnThrHisAspAspCysTyrGlyGluAlaGluLysLeuProAla 60

Qy         282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
               |||   :::::::::      ||||||   :::||||||||||||   |||:::
Db          61 CysAsnTyrMetMetSerGlyProTyrTyrAsnThrTyrSerTyrGluCysAsnGluGly 80

Qy         342 GCAATCACCTGT 353
                 :::||||||
Db          81 GluLeuThrCys 84
```

RESULT 12
PSNJ3K
phospholipase A2 (EC 3.1.1.4) III - monocled cobra
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Naja naja kaouthia, Naja naja siamensis (monocled cobra)
C;Date: 31-Mar-1992 #sequence_revision 14-Feb-1997 #text_change 16-Aug-2004
C;Accession: B00739; A00739
R;Joubert, F.J.; Taljaard, N.
Eur. J. Biochem. 112, 493-499, 1980
A;Title: Purification, some properties and amino-acid sequences of two phospholipases A
A;Reference number: A00739; MUID:81114211; PMID:7460933
A;Accession: B00739
A;Molecule type: protein
A;Residues: 1-70,'SC',73-119 <JOU>
A;Cross-references: UNIPARC:UPI0000172763
A;Note: the sequences in Table 9 and Figure 4 disagree with the sequence in Figure 5; th
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; metalloprotein; pres
F;4,67/Binding site: micellar substrate (Gln, Tyr) #status predicted
F;11-71,26-118,28-44,43-99,50-92,60-85,78-90/Disulfide bonds: #status predicted
F;27,29,31,48/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;47,93/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.:	4.14e-08	Length:	119
Score:	149.50	Matches:	28
Percent Similarity:	48.5%	Conservative:	5
Best Local Similarity:	41.2%	Mismatches:	12
Query Match:	11.9%	Indels:	23
DB:	1	Gaps:	2

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSNJ3K (1-119)

```
Qy         222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
               |||   ||||||||||||   ||||||||||||||||||:::|||:::|||:::
Db          40 LeuAspArgCysCysGlnValHisAspAsnCysTyrAspGluAlaGluLysIleSerGly 59

Qy         282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
               |||                |||||||      ||||||||||||   ||||||
Db          60 Cys--------------TrpProTyrPheLysThrTyrSerTyrGluCysSerGlnGly 74

Qy         342 GCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTC 401
                 :::||||||
Db          75 ThrLeuThrCysLys---------------------------------------------- 79

Qy         402 AGTAGGCCGGGGGGAAATAATAGT 425
                           ||||||||||||:::
Db          80 ---------GlyGlyAsnAsnAla 84
```

RESULT 13
PSNJAF
phospholipase A2 (EC 3.1.1.4) precursor [validated] - Chinese cobra
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Naja naja atra (Chinese cobra)
C;Date: 29-Jun-1981 #sequence_revision 31-Jan-1997 #text_change 16-Aug-2004
C;Accession: JC2137; I51017; A00740; S38512
R;Pan, F.M.; Yeh, M.S.; Chang, W.C.; Hung, C.C.; Chiou, S.H.
Biochem. Biophys. Res. Commun. 199, 969-976, 1994
A;Title: Sequence analysis and expression of phospholipase A2 from Taiwan cobra.
A;Reference number: JC2137; MUID:94183285; PMID:7510963
A;Accession: JC2137
A;Molecule type: mRNA
A;Residues: 1-146 <PAN1>
A;Cross-references: UNIPROT:P00598; UNIPARC:UPI0000131097; EMBL:X73225; NID:g395191; PI
A;Experimental source: venom gland
R;Pan, F.M.; Chang, W.C.; Chiou, S.H.
Biochem. Mol. Biol. Int. 33, 187-194, 1994
A;Title: cDNA and protein sequences coding for the precursor of phospholipase A2 from T.
A;Reference number: I51017; MUID:94362505; PMID:7521702
A;Accession: I51017
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-21,'S',23-75,'H',77-90,'S',92-146 <PAN2>
A;Cross-references: UNIPARC:UPI00001310B7; EMBL:X77755; NID:g456270; PIDN:CAA54802.1; PI
R;Tsai, I.H.; Wu, S.H.; Lo, T.B.
Toxicon 19, 141-152, 1981
A;Title: Complete amino acid sequence of a phospholipase A-2 from the venom of Naja naj
A;Reference number: A00740; MUID:81177953; PMID:7222082
A;Accession: A00740
A;Molecule type: protein
A;Residues: 28-110,'CA',113-133,'D',135,'D',137-139,'N',141-146,'E' <TSA>
A;Cross-references: UNIPARC:UPI0000172764
R;Chang, L.S.; Kuo, K.W.; Chang, C.C.
Biochim. Biophys. Acta 1202, 216-220, 1993
A;Title: Identification of functional involvement of tryptophan residues in phospholipa
A;Reference number: S38512; MUID:94002252; PMID:8399382
A;Accession: S38512
A;Molecule type: protein
A;Residues: 44-54;84-92 <CHA>
A;Cross-references: UNIPARC:UPI0000172765; UNIPARC:UPI0000172766
A;Note: sequence 1a in Table I has an extra Trp and an incorrect count as compared with
A;Note: Trp-45 is implicated in substrate binding
```
```

```
R;Scott, D.L.; Otwinowski, Z.; Sigler, P.B.
submitted to the Brookhaven Protein Data Bank, September 1992
A;Reference number: A51344; PDB:1POA
A;Contents: annotation; X-ray crystallography, 1.5 angstroms, residues 28-133,'D',135,'D
R;White, S.P.; Scott, D.L.; Otwinowski, Z.; Gelb, M.H.; Sigler, P.B.
Science 250, 1560-1563, 1990
A;Title: Crystal structure of cobra-venom phospholipase A2 in a complex with a transitio
A;Reference number: A58470; MUID:91111080; PMID:2274787
A;Contents: annotation; X-ray crystallography, 2.0 angstroms
C;Comment: Like most acidic phospholipases, this protein is not neurotoxic.
C;Genetics:
A;Gene: PLA2
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; metalloprotein; venc
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-146/Product: phospholipase A2 #status experimental <MAT>
F;31,94/Binding site: micellar substrate (Gln, Tyr) #status predicted
F;38-98,53-145,55-71,70-126,77-119,87-112,105-117/Disulfide bonds: #status experimental
F;54,56,58,75/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;74/Active site: His #status experimental
F;120/Active site: Asp #status predicted

Alignment Scores:
Pred. No.:                1.48e-07      Length:          146
Score:                    144.50        Matches:         27
Percent Similarity:       48.5%         Conservative:    6
Best Local Similarity:    39.7%         Mismatches:      12
Query Match:              11.5%         Indels:          23
DB:                       1             Gaps:            2

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSNJAF (1-146)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||   ||||||||||||||:::::::|||:::|||:::
Db         67 LeuAspArgCysCysGlnValHisAspAsnCysTyrAsnGluAlaGluLysIleSerGly 86

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||                 ||||||      ||||||||||||   ||||||
Db         87 Cys---------------TrpProTyrPheLysThrTyrSerTyrGluCysSerGlnGly 101

Qy        342 GCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTC 401
                 :::||||||
Db        102 ThrLeuThrCysLys---------------------------------------------- 106

Qy        402 AGTAGGCCGGGGGGAAATAATAGT 425
                        ||||||||||||:::
Db        107 ---------GlyGlyAsnAsnAla 111

RESULT 14
PSNJ2K
phospholipase A2 (EC 3.1.1.4) II - monocled cobra
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Naja naja kaouthia, Naja naja siamensis (monocled cobra)
C;Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 16-Aug-2004
C;Accession: A00739; A61011
R;Joubert, F.J.; Taljaard, N.
Eur. J. Biochem. 112, 493-499, 1980
A;Title: Purification, some properties and amino-acid sequences of two phospholipases A
A;Reference number: A00739; MUID:81114211; PMID:7460933
A;Accession: A00739
A;Molecule type: protein
A;Residues: 1-119 <JOU>
A;Cross-references: UNIPARC:UPI0000172767
R;Lo, C.H.; Chiou, S.H.
J. Chromatogr. 530, 129-136, 1990
A;Title: Simple fractionation of phospholipase A-2 analogues from snake venom by high-pe
A;Reference number: A61011; MUID:91116028; PMID:2277102
A;Accession: A61011
A;Molecule type: protein
A;Residues: 1-20 <LOA>
A;Cross-references: UNIPARC:UPI0000172768
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-s
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a micel
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; metalloprotein; presy
F;4,67/Binding site: micellar substrate (Gln, Tyr) #status predicted
F;11-71,26-118,28-44,43-99,50-92,60-85,78-90/Disulfide bonds: #status predicted
F;27,29,31,48/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;47,93/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:                1.94e-07      Length:          119
Score:                    143.50        Matches:         24
Percent Similarity:       63.0%         Conservative:    5
Best Local Similarity:    52.2%         Mismatches:      12
Query Match:              11.5%         Indels:          5
DB:                       1             Gaps:            1

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSNJ2K (1-119)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||   ||||||||||||||||||:::|||:::|||:::
Db         40 LeuAspArgCysCysGlnValHisAspAsnCysTyrAspGluAlaGluLysIleSerArg 59

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||                 ||||||      ||||||||||||   ||||||
Db         60 Cys---------------TrpProTyrPheLysThrTyrSerTyrGluCysSerGlnGly 74

Qy        342 GCAATCACCTGTAGCAGT 359
                 :::||||||   :::
Db         75 ThrLeuThrCysLysAsn 80

RESULT 15
PSOXG
phospholipase A2 (EC 3.1.1.4) taipoxin gamma chain - Australian taipan (tentative sequenc
N;Alternate names: phosphatidylcholine 2-acylhydrolase
C;Species: Oxyuranus scutellatus scutellatus (Australian taipan)
C;Date: 30-Nov-1980 #sequence_revision 30-Nov-1980 #text_change 16-Aug-2004
C;Accession: A00756
R;Fohlman, J.; Lind, P.; Eaker, D.
FEBS Lett. 84, 367-371, 1977
A;Title: Taipoxin, an extremely potent presynaptic snake venom neurotoxin. Elucidation o
A;Reference number: A00756; MUID:78084803; PMID:563806
A;Accession: A00756
A;Molecule type: protein
A;Residues: 1-133 <FOH>
A;Cross-references: UNIPROT:P00616; UNIPARC:UPI00001310D1
C;Comment: Taipoxin, a presynaptic neurotoxin of the venom, contains three noncovalently
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl-
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mice
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; glycoprotein; lipid degradation; metall
F;19-85,23-27,35-131,37-53,52-113,59-106,69-99,92-104/Disulfide bonds: #status predicted
F;36,38,40,57/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;56,107/Active site: His, Asp #status predicted
F;78/Binding site: carbohydrate (Asn) (covalent) #status experimental

Alignment Scores:
Pred. No.:                2.19e-07      Length:          133
Score:                    143.00        Matches:         27
Percent Similarity:       54.0%         Conservative:    7
Best Local Similarity:    42.9%         Mismatches:      19
Query Match:              11.4%         Indels:          10
DB:                       1             Gaps:            1

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PSOXG (1-133)

Qy        168 GGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCC 227
              |||   |||   ||||||                                        |||
Db         41 GlySerGlyThrProIleAsp----------------------------AspLeuAsp 50
```

```
Qy        228 AGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAA 287
              |||||||||:::|||||||||   ||||||   :::|||   ||||||   :::||||||
Db         51 ArgCysCysLysThrHisAspGluCysTyrAlaGluAlaGlyLysLeuSerAlaCysLys 70

Qy        288 TTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATC 347
                 :::|||      |||          ||||||||||||   |||:::         :::
Db         71 SerValLeuSerGluProAsnAsnAspThrTyrSerTyrGluCysAsnGluGlyGlnLeu 90

Qy        348 ACCTGTAGC 356
              ||||||:::
Db         91 ThrCysAsn 93
```

Search completed: February 8, 2006, 11:18:59
Job time : 25.9191 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31 ; Search time 32.3249 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-G7328_COPY_7000_7700
Perfect score: 1253
Sequence: 1 gtcgtgtcactgctgtccag.....tgcagcctcaacctcctgag 701

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_694_@runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	258	20.6	148	1	PA21B_HUMAN	P04054 homo sapien
2	222	17.7	146	1	PA21B_MOUSE	Q9z0y2 mus musculu
3	218	17.4	146	1	PA21B_CANFA	P06596 canis famil
4	216	17.2	144	1	PA2_RABIT	Q7m334 oryctolagus
5	216	17.2	146	1	PA21B_RAT	P04055 rattus norv
6	209	16.7	123	1	PA21B_SHEEP	P14419 ovis aries
7	209	16.7	145	1	PA21B_BOVIN	P00593 bos taurus
8	205	16.4	146	1	PA21B_PIG	P00592 sus scrofa
9	202	16.1	124	1	PA22_PIG	P04416 sus scrofa
10	201.5	16.1	132	1	PA21B_HORSE	P00594 equus cabal
11	188	15.0	146	1	PA21B_CAVPO	P43434 cavia porce
12	174	13.9	152	1	PA2I_LATSE	Q8jfb2 laticauda s
13	170	13.6	152	1	PA2H_LATSE	Q8jfg2 laticauda s
14	161	12.8	151	1	PA21B_OPHHA	P80966 ophiophagus
15	161	12.8	152	1	PA217_AUSSU	Q9pug7 austrelaps
16	159	12.7	119	1	PA22_ASPSC	P07037 aspidelaps
17	157	12.5	125	1	PA2E_NOTSC	Q9psn5 notechis sc
18	157	12.5	152	2	Q75S48_BUNCA	Q75s48 bungarus ca
19	152	12.1	154	1	PA21B_PSETE	Q9w7j4 pseudonaja
20	149.5	11.9	118	1	PA2H_BUNFA	P29601 bungarus fa
21	148	11.8	154	2	Q4VRI5_OXYSC	Q4vri5 oxyuranus s
22	145	11.6	152	1	PA216_AUSSU	Q9pug8 austrelaps
23	144.5	11.5	146	1	PA21_NAJAT	P00598 naja atra (
24	144.5	11.5	146	1	PA22_NAJKA	P00597 naja kaouth
25	144	11.5	154	1	PA22_PSETE	Q9w7j3 pseudonaja
26	143	11.4	133	1	PA23_OXYSC	P00616 oxyuranus s
27	143	11.4	152	2	Q4VRI6_OXYSC	Q4vri6 oxyuranus s
28	142.5	11.4	146	1	PA22_NAJAT	Q91133 naja atra (
29	142.5	11.4	146	1	PA2A_NAJSP	Q92084 naja sputat
30	142.5	11.4	146	1	PA2B_NAJSP	Q92085 naja sputat
31	142	11.3	152	1	PA23_OPHHA	Q9df56 ophiophagus
32	141.5	11.3	146	1	PA2C_NAJSP	Q92086 naja sputat
33	140.5	11.2	119	1	PA21_HEMHA	P00595 hemachatus
34	140.5	11.2	147	2	Q7T2Q4_9SAUR	Q7t2q4 bungarus fl
35	139.5	11.1	118	1	PA2X_BUNFA	P14411 bungarus fa
36	139.5	11.1	147	2	Q7T2Q5_9SAUR	Q7t2q5 bungarus fl
37	139	11.1	117	1	PA2A_PSEPO	P20258 pseudechis
38	138.5	11.1	118	1	PA2B_BUNFA	P00629 bungarus fa
39	138.5	11.1	119	1	PA2_NAJNA	P15445 naja naja (
40	138.5	11.1	120	2	Q65ZF5_NAJNA	Q65zf5 naja naja (
41	138.5	11.1	137	2	Q6SLM2_BUNCE	Q6slm2 bungarus ca
42	137.5	11.0	118	1	PA26_BUNFA	P00627 bungarus fa
43	137.5	11.0	118	1	PA2A_BUNFA	P00628 bungarus fa
44	137.5	11.0	126	1	PA22_NAJSG	P60044 naja sagitt
45	137.5	11.0	146	1	PA21_NAJKA	P00596 naja kaouth

ALIGNMENTS

RESULT 1
PA21B_HUMAN
ID PA21B_HUMAN STANDARD; PRT; 148 AA.
AC P04054;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE acylhydrolase) (Group IB phospholipase A2).
GN Name=PLA2G1B; Synonyms=PLA2, PLA2A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA].
RC TISSUE=Lung;
RX MEDLINE=87132925; PubMed=3028739;
RA Seilhamer J.J., Randall T.L., Yamanaka M., Johnson L.K.;
RT "Pancreatic phospholipase A2: isolation of the human gene and cDNAs
RT from porcine pancreas and human lung.";
RL DNA 5:519-527(1986).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT "NIEHS-SNPs, environmental genome project, NIEHS ES15478, Department
RT of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG Human chromosome 12 international sequencing consortium;
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP PROTEIN SEQUENCE OF 16-22.
RC TISSUE=Pancreas;
RX MEDLINE=82138816; PubMed=7060561;
RA Grataroli R., Dijkman R., Dutilh C.E., van der Ouderaa F.,

```
RA   de Haas G.H., Figarella C.;
RT   "Studies on prophospholipase A2 and its enzyme from human pancreatic
RT   juice. Catalytic properties and sequence of the N-terminal region.";
RL   Eur. J. Biochem. 122:111-117(1982).
RN   [5]
RP   PROTEIN SEQUENCE OF 23-148.
RC   TISSUE=Pancreas;
RX   MEDLINE=83283533; PubMed=6349696; DOI=10.1016/0167-4838(83)90126-7;
RA   Verheij H.M., Westerman J., Sternby B., de Haas G.H.;
RT   "The complete primary structure of phospholipase A2 from human
RT   pancreas.";
RL   Biochim. Biophys. Acta 747:93-99(1983).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; M21056; AAA60107.1; -; Genomic_DNA.
DR   EMBL; M22970; AAA60107.1; JOINED; Genomic_DNA.
DR   EMBL; M21054; AAA36450.1; -; mRNA.
DR   EMBL; AY438977; AAR05441.1; -; Genomic_DNA.
DR   EMBL; AC003982; AAB95635.1; -; Genomic_DNA.
DR   PIR; C25793; PSHU.
DR   HSSP; P00592; 1HN4.
DR   Ensembl; ENSG00000170890; Homo sapiens.
DR   HGNC; HGNC:9030; PLA2G1B.
DR   MIM; 172410; -.
DR   GO; GO:0004623; F:phospholipase A2 activity; TAS.
DR   GO; GO:0007015; P:actin filament organization; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW   Metal-binding; Polymorphism; Signal.
FT   SIGNAL        1     15
FT   PROPEP       16     22       Activation peptide.
FT   CHAIN        23    148       Phospholipase A2.
FT   ACT_SITE     70     70       By similarity.
FT   ACT_SITE    121    121       By similarity.
FT   METAL        50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        71     71       Calcium (By similarity).
FT   DISULFID     33     99
FT   DISULFID     49    146
FT   DISULFID     51     67
FT   DISULFID     66    127
FT   DISULFID     73    120
FT   DISULFID     83    113
FT   DISULFID    106    118
FT   VARIANT      16     16       D -> A (in dbSNP:5632).
FT                                /FTId=VAR_011911.
FT   VARIANT      89     89       N -> K (in dbSNP:5636).
FT                                /FTId=VAR_011913.
FT   VARIANT      89     89       N -> T (in dbSNP:5635).
FT                                /FTId=VAR_011912.
FT   CONFLICT    144    144       Missing (in Ref. 5).
SQ   SEQUENCE   148 AA;  16360 MW;  C8E38B2B64AEE8CB CRC64;
```

```
Alignment Scores:
Pred. No.:              4.08e-20        Length:         148
Score:                  258.00          Matches:        55
Percent Similarity:     59.8%           Conservative:   6
Best Local Similarity:  53.9%           Mismatches:     12
Query Match:            20.6%           Indels:         29
DB:                     1               Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_HUMAN (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
              |||   |||                     :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
              |||:::                                          |||   :::||||||||
Db         58 rProValAsp-----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109
```

```
RESULT 2
PA21B_MOUSE
ID   PA21B_MOUSE    STANDARD;      PRT;   146 AA.
AC   Q9Z0Y2; Q9D7E2; Q9D884;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2) (PLA2-Ib).
GN   Name=Pla2g1b; Synonyms=Pla2;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=129/SvJ;
RA   Mandal A.K., Zhang Z., Mukherjee A.B.;
RT   "Isolation, characterization and chromosomal localization of mouse
RT   sPLA2-Ib gene.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=99167464; PubMed=10066760; DOI=10.1074/jbc.274.11.7043;
RA   Cupillard L., Mulherkar R., Gomez N., Kadam S., Valentin E.,
RA   Lazdunski M., Lambeau G.;
RT   "Both group IB and group IIA secreted phospholipases A2 are natural
RT   ligands of the mouse 180-kDa M-type receptor.";
RL   J. Biol. Chem. 274:7043-7051(1999).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=BALB/c; TISSUE=Lung;
RX   MEDLINE=20002676; PubMed=10531350; DOI=10.1074/jbc.274.44.31476;
RA   Bingham C.O. III, Fijneman R.J.A., Friend D.S., Goddeau R.P.,
```

```
RA   Rogers R.A., Austen K.F., Arm J.P.;
RT   "Low molecular weight group IIA and group V phospholipase A(2) enzymes
RT   have different intracellular locations in mouse bone marrow-derived
RT   mast cells.";
RL   J. Biol. Chem. 274:31476-31484(1999).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=20156393; PubMed=10689188; DOI=10.1016/S0378-1119(00)00006-8;
RA   Richmond B.L., Hui D.Y.;
RT   "Molecular structure and tissue-specific expression of the mouse
RT   pancreatic phospholipase A2 gene.";
RL   Gene 244:65-72(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=129/SvJ; TISSUE=Pancreas;
RA   Mandal A.K., Zhang Z., Popescu N., Mukherjee A.B.;
RT   "Role of sPLA2-I in colorectal tumorigenesis.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   STRAIN=C57BL/6J;
RC   TISSUE=Pancreas, Small intestine, Stomach, and Tongue;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AF094611; AAF02298.1; -; Genomic_DNA.
DR   EMBL; AF097637; AAD19896.1; -; mRNA.
DR   EMBL; AF162712; AAD45806.1; -; mRNA.
DR   EMBL; AF187852; AAF44297.1; -; mRNA.
DR   EMBL; AF094610; AAG27064.1; -; mRNA.
DR   EMBL; AK028104; BAC25749.1; -; mRNA.
DR   EMBL; AK028134; BAC25763.1; -; mRNA.
DR   EMBL; AK008936; BAB25978.1; -; mRNA.
DR   EMBL; AK007730; BAB25218.1; -; mRNA.
DR   EMBL; AK007797; BAB25263.1; -; mRNA.
DR   EMBL; AK008664; BAB25819.1; -; mRNA.
DR   EMBL; AK008668; BAB25822.1; -; mRNA.
DR   EMBL; AK008841; BAB25922.1; -; mRNA.
DR   EMBL; AK008934; BAB25976.1; -; mRNA.
DR   EMBL; AK008331; BAB25608.1; -; mRNA.
DR   EMBL; AK009314; BAB26212.1; -; mRNA.
DR   HSSP; P00593; 1GH4.
DR   Ensembl; ENSMUSG00000029522; Mus musculus.
DR   MGI; MGI:101842; Pla2g1b.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0004623; F:phospholipase A2 activity; IDA.
DR   GO; GO:0005102; F:receptor binding; IDA.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0009395; P:phospholipid catabolism; TAS.
DR   GO; GO:0006950; P:response to stress; TAS.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
FT   SIGNAL       1     15       By similarity.
FT   PROPEP      16     22       Activation peptide (By similarity).
FT   CHAIN       23    146       Phospholipase A2.
FT   ACT_SITE    70     70       By similarity.
FT   ACT_SITE   121    121       By similarity.
FT   METAL        50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        71     71       Calcium (By similarity).
FT   DISULFID     33     99       By similarity.
FT   DISULFID     49    146       By similarity.
FT   DISULFID     51     67       By similarity.
FT   DISULFID     66    127       By similarity.
FT   DISULFID     73    120       By similarity.
FT   DISULFID     83    113       By similarity.
FT   DISULFID    106    118       By similarity.
FT   CONFLICT     52     52       G -> S (in Ref. 6; BAB26212).
FT   CONFLICT     58     58       T -> P (in Ref. 6; BAB26212).
FT   CONFLICT     78     79       KK -> EN (in Ref. 6; BAB26212).
FT   CONFLICT     87     87       I -> R (in Ref. 6; BAB26212).
FT   CONFLICT     95     95       Y -> F (in Ref. 6; BAB26212).
FT   CONFLICT    102    102       S -> G (in Ref. 6; BAB26212).
FT   CONFLICT    122    122       R -> S (in Ref. 6; BAB25608).
SQ   SEQUENCE   146 AA;  16290 MW;  59500C68845B7C81 CRC64;

Alignment Scores:
Pred. No.:              5.36e-16        Length:         146
Score:                  222.00          Matches:        37
Percent Similarity:     91.5%           Conservative:   6
Best Local Similarity:  78.7%           Mismatches:     4
Query Match:            17.7%           Indels:         0
DB:                     1               Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_MOUSE (1-146)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||  ||||||||||||||||||||:::||||||  ||||||||||||||:::|||
Db         63 LeuAspArgCysCysGlnThrHisAspHisCysTyrSerGlnAlaLysLysLeuGluSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||||||||||||:::|||  |||||||||:::||||||||||||||||||||||||||||
Db         83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrSerCysSerGlySer 102
```

```
Qy        342 GCAATCACCTGTAGCAGTAGG 362
                ||||||||||||::::::
Db        103 GluIleThrCysSerAlaLys 109

RESULT 3
PA21B_CANFA
ID   PA21B_CANFA    STANDARD;      PRT;   146 AA.
AC   P06596;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC   Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=86223862; PubMed=3754861;
RA   Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
RA   Teraoka H., Okamoto M.;
RT   "Dog and rat pancreatic phospholipases A2: complete amino acid
RT   sequences deduced from complementary DNAs.";
RL   J. Biochem. 99:733-739(1986).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=90306027; PubMed=2142076;
RA   Kerfelec B., Laforge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
RT   "Isolation and sequence of the canine pancreatic phospholipase A2
RT   gene.";
RL   Eur. J. Biochem. 190:299-304(1990).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=87175472; PubMed=3562437;
RA   Kerfelec B., Laforge K.S., Puigserver A., Scheele G.A.;
RT   "Primary structures of canine pancreatic lipase and phospholipase A2
RT   messenger RNAs.";
RL   Pancreas 1:430-437(1986).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; D00035; BAA00023.1; -; mRNA.
DR   EMBL; M35301; AAA30883.1; -; mRNA.
DR   PIR; S11316; PSDG.
DR   HSSP; P00592; 1HN4.
DR   Ensembl; ENSCAFG00000010263; Canis familiaris.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
FT   SIGNAL        1     15
FT   PROPEP       16     22
FT   CHAIN        23    146       Phospholipase A2.
FT   ACT_SITE     70     70       By similarity.
FT   ACT_SITE    121    121       By similarity.
FT   METAL        50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        71     71       Calcium (By similarity).
FT   DISULFID     33     99       By similarity.
FT   DISULFID     49    146       By similarity.
FT   DISULFID     51     67       By similarity.
FT   DISULFID     66    127       By similarity.
FT   DISULFID     73    120       By similarity.
FT   DISULFID     83    113       By similarity.
FT   DISULFID    106    118       By similarity.
SQ   SEQUENCE   146 AA;  16236 MW;  F6258ED9527F3692 CRC64;

Alignment Scores:
Pred. No.:              1.54e-15      Length:         146
Score:                  218.00        Matches:        48
Percent Similarity:     54.9%         Conservative:   8
Best Local Similarity:  47.1%         Mismatches:     17
Query Match:            17.4%         Indels:         29
DB:                     1             Gaps:           3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_CANFA (1-146)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                 |||
Db         19 IleSerProArgAlaValTrpGlnPheArgAsnMetIleLysCysThrIleProGluSer 38

Qy        140 CATCCC--------------------CACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTG 178
              |||                  :::      :::||| |||   |||
Db         39 AspProLeuLysAspTyrAsnAspTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
               |||:::                                    |||   :::||||||||
Db         58 rProValAsp----------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||:::||||||   :::||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspHisCysTyrSerGluAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||       ||||||||||||||||||||||||   ||||||||||||||
Db         88 pAsnProTyrThrLysIleTyrSerTyrSerCysSerGlySerGluIleThrCysSerSe 108

Qy        359 TAGG 362
              |:::
Db        108 rLys 109

RESULT 4
PA2_RABIT
ID   PA2_RABIT      STANDARD;      PRT;   144 AA.
AC   Q7M334;
DT   05-JUL-2004 (Rel. 44, Created)
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae;
OC   Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pancreas;
RX   MEDLINE=93249443; PubMed=7916604;
RA   Kumar V.B.;
RT   "Cloning and expression of rabbit pancreatic phospholipase A2.";
RL   Biochem. Biophys. Res. Commun. 192:683-692(1993).
```

```
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides (By similarity).
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   PIR; JN0480; JN0480.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
FT   SIGNAL       1     15       By similarity.
FT   PROPEP      16     22       Activation peptide.
FT   CHAIN       23    144       Phospholipase A2.
FT   ACT_SITE    70     70       By similarity.
FT   ACT_SITE   121    121       By similarity.
FT   METAL        50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        71     71       Calcium (By similarity).
FT   DISULFID     33     99       By similarity.
FT   DISULFID     49    144       By similarity.
FT   DISULFID     51     67       By similarity.
FT   DISULFID     66    127       By similarity.
FT   DISULFID     73    120       By similarity.
FT   DISULFID     83    113       By similarity.
FT   DISULFID    106    118       By similarity.
SQ   SEQUENCE   144 AA;  15733 MW;  DF404C1F28AE932A CRC64;

Alignment Scores:
Pred. No.:                2.6e-15        Length:          144
Score:                    216.00         Matches:         45
Percent Similarity:       55.9%          Conservative:    12
Best Local Similarity:    44.1%          Mismatches:      16
Query Match:              17.2%          Indels:          29
DB:                       1              Gaps:            3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA2_RABIT (1-144)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              :::||||||   |||   |||:::                                 |||
Db         19 ValSerProThrAlaLeuTrpGlnPheArgGlyMetIleGlnCysThrIleProGlySer  38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                 |||:::|||                    :::     :::||| |||   |||
Db         39 SerProTyrLeuGluPheAsnGlyTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh  58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
               |||:::                                  |||   ||||||||||
Db         58 rProValAsp-----------------------------GluLeuAspArgCysCysGl  68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||   ||||||   ||||||||||||||   ||||||   ||||||:::||
Db         68 nThrHisAspGlnCysTyrThrGlnAlaLysLysLeuSerSerCysSerPheLeuValAs  88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||::::::||||||||||||||||||:::   :::||||||||||
Db         88 pAsnProTyrThrAsnSerTyrSerTyrSerCysSerGlyThrThrValThrCysSerSe 108

Qy        359 TAGG 362
                |:::
Db        108 rLys 109

RESULT 5
PA21B_RAT
ID   PA21B_RAT      STANDARD;      PRT;   146 AA.
AC   P04055;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2).
GN   Name=Pla2g1b;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=86223862; PubMed=3754861;
RA   Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
RA   Teraoka H., Okamoto M.;
RT   "Dog and rat pancreatic phospholipases A2: complete amino acid
RT   sequences deduced from complementary DNAs.";
RL   J. Biochem. 99:733-739(1986).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=89076874; PubMed=2909239; DOI=10.1016/0167-4781(89)90141-3;
RA   Sakata T., Nakamura E., Tsuruta Y., Tamaki M., Teraoka H., Tojo H.,
RA   Ono T., Okamoto M.;
RT   "Presence of pancreatic-type phospholipase A2 mRNA in rat gastric
RT   mucosa and lung.";
RL   Biochim. Biophys. Acta 1007:124-126(1989).
RN   [3]
RP   PROTEIN SEQUENCE OF 23-54.
RX   MEDLINE=85054750; PubMed=6501264;
RA   Ono T., Tojo H., Inoue K., Kagamiyama H., Yamano T., Okamoto M.;
RT   "Rat pancreatic phospholipase A2: purification, characterization, and
RT   N-terminal amino acid sequence.";
RL   J. Biochem. 96:785-792(1984).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; D00036; BAA00024.1; -; mRNA.
DR   PIR; A92008; PSRT.
DR   HSSP; P00593; 4BP2.
DR   Ensembl; ENSRNOG00000001153; Rattus norvegicus.
DR   RGD; 61949; Pla2g1b.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW   Metal-binding; Signal.
FT   SIGNAL       1     15
```

```
FT   PROPEP       16     22       Activation peptide.
FT   CHAIN        23    146       Phospholipase A2.
FT   ACT_SITE     70     70
FT   ACT_SITE    121    121
FT   METAL        50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        71     71       Calcium (By similarity).
FT   DISULFID     33     99       By similarity.
FT   DISULFID     49    146       By similarity.
FT   DISULFID     51     67       By similarity.
FT   DISULFID     66    127       By similarity.
FT   DISULFID     73    120       By similarity.
FT   DISULFID     83    113       By similarity.
FT   DISULFID    106    118       By similarity.
SQ   SEQUENCE   146 AA;  16424 MW;  7EC4F7A491B913D0 CRC64;

Alignment Scores:
Pred. No.:                2.61e-15       Length:        146
Score:                    216.00         Matches:       35
Percent Similarity:       89.4%          Conservative:  7
Best Local Similarity:    74.5%          Mismatches:    5
Query Match:              17.2%          Indels:        0
DB:                       1              Gaps:          0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_RAT (1-146)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||||||||||||||:::||||||:::||||||||||||||||:::|||
Db         63 LeuAspArgCysCysGlnThrHisAspHisCysTyrAsnGlnAlaLysLysLeuGluSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||||||||:::|||   |||||||||:::||||||||||||   |||||||||:::
Db         83 CysLysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrLysCysSerGlyAsn 102

Qy        342 GCAATCACCTGTAGCAGTAGG 362
                  ||||||||||||   :::
Db        103 ValIleThrCysSerAspLys 109

RESULT 6
PA21B_SHEEP
ID   PA21B_SHEEP    STANDARD;      PRT;   123 AA.
AC   P14419;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase)
DE   (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC   Pecora; Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   PROTEIN SEQUENCE.
RC   TISSUE=Pancreas;
RA   Verheij H.M., Raykova D., Dijkman R., Lenting H.B.M., de Haas G.H.;
RT   "The primary structure and some kinetic properties of phospholipase A2
RT   from sheep pancreas.";
RL   Recl. Trav. Chim. Pays-Bas 107:203-207(1988).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   HSSP; P00593; 4BP2.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW   Metal-binding.
FT   ACT_SITE     48     48       By similarity.
FT   ACT_SITE     99     99       By similarity.
FT   METAL        28     28       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        30     30       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        32     32       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        49     49       Calcium (By similarity).
FT   DISULFID     11     77       By similarity.
FT   DISULFID     27    123       By similarity.
FT   DISULFID     29     45       By similarity.
FT   DISULFID     44    105       By similarity.
FT   DISULFID     51     98       By similarity.
FT   DISULFID     61     91       By similarity.
FT   DISULFID     84     96       By similarity.
SQ   SEQUENCE   123 AA;  13817 MW;  F69878C15DF6C4F9 CRC64;

Alignment Scores:
Pred. No.:                1.6e-14        Length:        123
Score:                    209.00         Matches:       36
Percent Similarity:       84.8%          Conservative:  3
Best Local Similarity:    78.3%          Mismatches:    7
Query Match:              16.7%          Indels:        0
DB:                       1              Gaps:          0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_SHEEP (1-123)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   |||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         41 LeuAspArgCysCysGlnThrHisAspAsnCysTyrLysGlnAlaLysLysLeuAspSer 60

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||   |||:::|||   |||||||||::::::||||||||||||||||||||||
Db         61 CysLysValLeuValAspAsnProTyrThrAsnSerTyrSerTyrSerCysSerAsnLys 80

Qy        342 GCAATCACCTGTAGCAGT 359
                  ||||||||||||||||
Db         81 GlnIleThrCysSerSer 86

RESULT 7
PA21B_BOVIN
ID   PA21B_BOVIN    STANDARD;      PRT;   145 AA.
AC   P00593;
DT   21-JUL-1986 (Rel. 01, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC   Pecora; Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
```

```
RX   MEDLINE=87174835; PubMed=3562249;
RA   Tanaka T., Kimura S., Ota Y.;
RT   "Sequence of a cDNA coding for bovine pancreatic phospholipase A2.";
RL   Nucleic Acids Res. 15:3178-3178(1987).
RN   [2]
RP   PROTEIN SEQUENCE OF 16-22.
RC   TISSUE=Pancreas;
RA   Dutilh C.E., van Doren P.J., Verheul F.E.A.M., de Haas G.H.;
RT   "Isolation and properties of prophospholipase A2 from ox and sheep
RT   pancreas.";
RL   Eur. J. Biochem. 53:91-97(1975).
RN   [3]
RP   PROTEIN SEQUENCE OF 23-145.
RX   MEDLINE=78084411; PubMed=620674;
RA   Fleer E.A.M., Verheij H.M., de Haas G.H.;
RT   "The primary structure of bovine pancreatic phospholipase A2.";
RL   Eur. J. Biochem. 82:261-269(1978).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS), AND DISULFIDE BONDS.
RX   MEDLINE=79050564; PubMed=712836;
RA   Dijkstra B.W., Drenth J., Kalk K.H., Vandermaelen P.J.;
RT   "Three-dimensional structure and disulfide bond connections in bovine
RT   pancreatic phospholipase A2.";
RL   J. Mol. Biol. 124:53-60(1978).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS), ACTIVE SITE, AND CATALYTIC
RP   MECHANISM.
RX   MEDLINE=81123082; PubMed=7464926;
RA   Dijkstra B.W., Drenth J., Kalk K.H.;
RT   "Active site and catalytic mechanism of phospholipase A2.";
RL   Nature 289:604-606(1981).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX   MEDLINE=81267392; PubMed=7265241;
RA   Dijkstra B.W., Kalk K.H., Hol W.G.J., Drenth J.;
RT   "Structure of bovine pancreatic phospholipase A2 at 1.7A resolution.";
RL   J. Mol. Biol. 147:97-123(1981).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF PROENZYME.
RA   Dijkstra B.W., van Nes G.J.H., Kalk K.H., Brandenburg N.P.,
RA   Hol W.G.J., Drenth J.;
RT   "The structure of bovine pancreatic prophospholipase A2 at 3.0-A
RT   resolution.";
RL   Acta Crystallogr. B 38:793-799(1982).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.72 ANGSTROMS).
RX   MEDLINE=97228082; PubMed=9115986; DOI=10.1021/bi961576x;
RA   Sekar K., Yu B.Z., Rogers J., Lutton J., Liu X., Chen X., Tsai M.-D.,
RA   Jain M.K., Sundaralingam M.;
RT   "Phospholipase A2 engineering. Structural and functional roles of the
RT   highly conserved active site residue aspartate-99.";
RL   Biochemistry 36:3104-3114(1997).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=98039151; PubMed=9369492; DOI=10.1021/bi971370b;
RA   Sekar K., Eswaramoorthy S., Jain M.K., Sundaralingam M.;
RT   "Crystal structure of the complex of bovine pancreatic phospholipase
RT   A2 with the inhibitor 1-hexadecyl-3-(trifluoroethyl)-sn-glycero-2-
RT   phosphomethanol.";
RL   Biochemistry 36:14186-14191(1997).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS) OF MUTANTS.
RX   MEDLINE=99190898; PubMed=10089353; DOI=10.1107/S0907444998013699;
RA   Sekar K., Biswas R., Li Y., Tsai M., Sundaralingam M.;
RT   "Structures of the catalytic site mutants D99A and H48Q and the
RT   calcium-loop mutant D49E of phospholipase A2.";
RL   Acta Crystallogr. D 55:443-447(1999).
RN   [11]
RP   STRUCTURE BY NMR.
RX   MEDLINE=99175086; PubMed=10074343; DOI=10.1021/bi982211a;
RA   Yuan C., Byeon I.-J.L., Li Y., Tsai M.-D.;
RT   "Structural analysis of phospholipase A2 from functional perspective.
RT   1. Functionally relevant solution structure and roles of the hydrogen-
RT   bonding network.";
RL   Biochemistry 38:2909-2918(1999).
RN   [12]
RP   STRUCTURE BY NMR OF MUTANTS.
RX   MEDLINE=99175087; PubMed=10074344; DOI=10.1021/bi9822123;
RA   Yuan C., Byeon I.-J.L., Poi M.-J., Tsai M.-D.;
RT   "Structural analysis of phospholipase A2 from functional perspective.
RT   2. Characterization of a molten globule-like state induced by site-
RT   specific mutagenesis.";
RL   Biochemistry 38:2919-2929(1999).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ----------------------------------------------------------------------
DR   EMBL; Y00120; CAA68303.1; -; mRNA.
DR   PIR; A27508; PSBOA.
DR   PDB; 1BP2; X-ray; @=23-145.
DR   PDB; 1BPQ; X-ray; @=23-145.
DR   PDB; 1BVM; NMR; A=23-145.
DR   PDB; 1C74; X-ray; A=23-145.
DR   PDB; 1CEH; X-ray; @=23-145.
DR   PDB; 1FDK; X-ray; @=23-145.
DR   PDB; 1G4I; X-ray; A=23-145.
DR   PDB; 1GH4; X-ray; A=23-141.
DR   PDB; 1IRB; X-ray; @=23-141.
DR   PDB; 1KVW; X-ray; @=23-145.
DR   PDB; 1KVX; X-ray; @=23-145.
DR   PDB; 1KVY; X-ray; @=23-145.
DR   PDB; 1MKS; X-ray; @=23-145.
DR   PDB; 1MKT; X-ray; @=23-145.
DR   PDB; 1MKU; X-ray; @=23-145.
DR   PDB; 1MKV; X-ray; @=23-145.
DR   PDB; 1O2E; X-ray; A=23-145.
DR   PDB; 1O3W; X-ray; A=23-145.
DR   PDB; 1UNE; X-ray; @=23-145.
DR   PDB; 1VKQ; X-ray; A=23-145.
DR   PDB; 1VL9; X-ray; A=23-144.
DR   PDB; 2BP2; X-ray; @=17-145.
DR   PDB; 2BPP; X-ray; @=23-145.
DR   PDB; 3BP2; X-ray; @=24-145.
DR   PDB; 4BP2; X-ray; @=16-145.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   3D-structure; Calcium; Direct protein sequencing; Hydrolase;
KW   Lipid degradation; Metal-binding; Pyrrolidone carboxylic acid; Signal.
FT   SIGNAL       1     15
FT   PROPEP      16     22       Removed by trypsin.
FT   CHAIN       23    145       Phospholipase A2.
FT   ACT_SITE    70     70
FT   ACT_SITE   121    121
FT   METAL       50     50       Calcium (via carbonyl oxygen).
FT   METAL       52     52       Calcium (via carbonyl oxygen).
FT   METAL       54     54       Calcium (via carbonyl oxygen).
FT   METAL       71     71       Calcium.
FT   MOD_RES     16     16       Pyrrolidone carboxylic acid.
FT   DISULFID    33     99
FT   DISULFID    49    145
```

```
FT   DISULFID     51     67
FT   DISULFID     66    127
FT   DISULFID     73    120
FT   DISULFID     83    113
FT   DISULFID    106    118
FT   CONFLICT    144    144       N -> K (in Ref. 1).
FT   HELIX         24     34
FT   TURN          36     37
FT   HELIX         40     43
FT   TURN          44     44
FT   STRAND        47     47
FT   TURN          48     50
FT   STRAND        51     51
FT   HELIX         62     79
FT   HELIX         81     85
FT   TURN          86     87
FT   HELIX         90     92
FT   STRAND        97    100
FT   TURN         101    102
FT   STRAND       103    106
FT   TURN         108    109
FT   HELIX        112    129
FT   TURN         130    130
FT   HELIX        135    137
FT   STRAND       138    138
FT   TURN         139    139
FT   HELIX        142    145
SQ   SEQUENCE   145 AA;  16002 MW;  17A1E04B0C22F668 CRC64;

Alignment Scores:
Pred. No.:              1.65e-14       Length:         145
Score:                  209.00         Matches:        36
Percent Similarity:     84.8%          Conservative:   3
Best Local Similarity:  78.3%          Mismatches:     7
Query Match:            16.7%          Indels:         0
DB:                     1              Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_BOVIN (1-145)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db         63 LeuAspArgCysCysGlnThrHisAspAsnCysTyrLysGlnAlaLysLysLeuAspSer 82

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||   |||:::|||   |||||||||:::   ||||||||||||||||||   :::
Db         83 CysLysValLeuValAspAsnProTyrThrAsnAsnTyrSerTyrSerCysSerAsnAsn 102

Qy        342 GCAATCACCTGTAGCAGT 359
                |||||||||||||||
Db        103 GluIleThrCysSerSer 108

RESULT 8
PA21B_PIG
ID   PA21B_PIG      STANDARD;      PRT;   146 AA.
AC   P00592;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2, major isoenzyme precursor (EC 3.1.1.4)
DE   (Phosphatidylcholine 2-acylhydrolase) (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC   Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pancreas;
RX   MEDLINE=87231066; PubMed=3295782;
RA   de Geus P., van den Bergh C.J., Kuipers O., Verheij H.M.,
RA   Hoekstra W.P.M., de Haas G.H.;
RT   "Expression of porcine pancreatic phospholipase A2. Generation of
RT   active enzyme by sequence-specific cleavage of a hybrid protein from
RT   Escherichia coli.";
RL   Nucleic Acids Res. 15:3743-3759(1987).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pancreas;
RX   MEDLINE=87132925; PubMed=3028739;
RA   Seilhamer J.J., Randall T.L., Yamanaka M., Johnson L.K.;
RT   "Pancreatic phospholipase A2: isolation of the human gene and cDNAs
RT   from porcine pancreas and human lung.";
RL   DNA 5:519-527(1986).
RN   [3]
RP   PROTEIN SEQUENCE OF 16-146.
RX   MEDLINE=71014235; PubMed=5528841;
RA   de Haas G.H., Slotboom A.J., Bonsen P.P.M., van Deenen L.L.M.,
RA   Maroux S., Puigserver A., Desnuelle P.;
RT   "Studies on phospholipase A and its zymogen from porcine pancreas. I.
RT   The complete amino acid sequence.";
RL   Biochim. Biophys. Acta 221:31-53(1970).
RN   [4]
RP   SEQUENCE REVISION.
RX   MEDLINE=77222092; PubMed=884127; DOI=10.1016/0005-2795(77)90076-9;
RA   Puijk W.C., Verheij H.M., de Haas G.H.;
RT   "The primary structure of phospholipase A2 from porcine pancreas. A
RT   reinvestigation.";
RL   Biochim. Biophys. Acta 492:254-259(1977).
RN   [5]
RP   DISULFIDE BONDS.
RX   MEDLINE=71014236; PubMed=4919729;
RA   de Haas G.H., Slotboom A.J., Bonsen P.P.M., Nieuwenhuizen W.,
RA   van Deenen L.L.M., Maroux S., Dlouha V., Desnuelle P.;
RT   "Studies on phospholipase A and its zymogen from porcine pancreas. II.
RT   The assignment of the position of the six disulfide bridges.";
RL   Biochim. Biophys. Acta 221:54-61(1970).
RN   [6]
RP   ACYLATION OF LYS-78.
RC   TISSUE=Pancreas;
RX   MEDLINE=89255488; PubMed=2498336;
RA   Tomasselli A.G., Hui J., Fisher J., Zuercher-Neely H., Reardon H.M.,
RA   Oriaku E., Kezdy F.J., Heinrikson R.L.;
RT   "Dimerization and activation of porcine pancreatic phospholipase A2
RT   via substrate level acylation of lysine 56.";
RL   J. Biol. Chem. 264:10041-10047(1989).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND ACTIVE AND BINDING SITES.
RX   MEDLINE=83268704; PubMed=6876174;
RA   Dijkstra B.W., Renetseder R., Kalk K.H., Hol W.G.J., Drenth J.;
RT   "Structure of porcine pancreatic phospholipase A2 at 2.6-A resolution
RT   and comparison with bovine phospholipase A2.";
RL   J. Mol. Biol. 168:163-179(1983).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=91073405; PubMed=2254938;
RA   Thunnissen M.M.G.M., Kalk K.H., Drenth J., Dijkstra B.W.;
RT   "Structure of an engineered porcine phospholipase A2 with enhanced
RT   activity at 2.1-A resolution. Comparison with the wild-type porcine
RT   and Crotalus atrox phospholipase A2.";
RL   J. Mol. Biol. 216:425-439(1990).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX   MEDLINE=91015391; PubMed=2215698; DOI=10.1038/347689a0;
RA   Thunnissen M.M.G.M., Ab E., Kalk K.H., Drenth J., Dijkstra B.W.,
RA   Kuipers O.P., Dijkman R., de Haas G.H., Verheij H.M.;
RT   "X-ray structure of phospholipase A2 complexed with a substrate-
RT   derived inhibitor.";
RL   Nature 347:689-691(1990).
RN   [10]
RP   STRUCTURE BY NMR.
RX   MEDLINE=91175768; PubMed=2007145;
RA   Dekker N., Peters A.R., Slotboom A.J., Boelens R., Kaptein R.,
RA   de Haas G.H.;
RT   "Porcine pancreatic phospholipase A2: sequence-specific 1H and 15N NMR
RT   assignments and secondary structure.";
```

```
RL   Biochemistry 30:3135-3147(1991).
RN   [11]
RP   STRUCTURE BY NMR.
RX   MEDLINE=95393214; PubMed=7664098;
RA   van den Berg B., Tessari M., Boelens R., Dijkman R., de Haas G.H.,
RA   Kaptein R., Verheij H.M.;
RT   "NMR structures of phospholipase A2 reveal conformational changes
RT   during interfacial activation.";
RL   Nat. Struct. Biol. 2:402-406(1995).
RN   [12]
RP   STRUCTURE BY NMR.
RX   MEDLINE=96016184; PubMed=7556053;
RA   van den Berg B., Tessari M., de Haas G.H., Verheij H.M., Boelens R.,
RA   Kaptein R.;
RT   "Solution structure of porcine pancreatic phospholipase A2.";
RL   EMBO J. 14:4123-4131(1995).
RN   [13]
RP   STRUCTURE BY NMR.
RX   MEDLINE=95218441; PubMed=7703697;
RA   van den Berg B., Tessari M., Boelens R., Dijkman R., Kaptein R.,
RA   de Haas G.H., Verheij H.M.;
RT   "Solution structure of porcine pancreatic phospholipase A2 complexed
RT   with micelles and a competitive inhibitor.";
RL   J. Biomol. NMR 5:110-121(1995).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit.
CC   -!- SUBUNIT: Monomer or homodimer.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: Acylation causes dimerization.
CC   -!- MISCELLANEOUS: Loss of activity upon alkylation of His-70 with p-
CC       bromo phenacyl bromide; Ca(2+) and Ba(2+) protect against
CC       inactivation.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; Y00146; CAA68341.1; -; mRNA.
DR   EMBL; M21055; AAA31101.1; -; mRNA.
DR   PIR; B25793; PSPGA.
DR   PDB; 1FX9; X-ray; A/B=23-146.
DR   PDB; 1FXF; X-ray; A/B=23-146.
DR   PDB; 1HN4; X-ray; A/B=16-146.
DR   PDB; 1L8S; X-ray; A/B=23-146.
DR   PDB; 1P2P; X-ray; @=23-146.
DR   PDB; 1PIR; NMR; @=23-146.
DR   PDB; 1PIS; NMR; @=23-146.
DR   PDB; 1SFV; NMR; @=23-146.
DR   PDB; 1SFW; NMR; @=23-146.
DR   PDB; 1Y6O; X-ray; A/B=23-146.
DR   PDB; 1Y6P; X-ray; A/B=23-146.
DR   PDB; 2PHI; X-ray; A/B=23-146.
DR   PDB; 3P2P; X-ray; A/B=23-146.
DR   PDB; 4P2P; X-ray; @=23-146.
DR   PDB; 5P2P; X-ray; A/B=24-146.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   3D-structure; Calcium; Direct protein sequencing; Hydrolase;
KW   Lipid degradation; Lipoprotein; Metal-binding; Palmitate;
KW   Pyrrolidone carboxylic acid; Signal.
FT   SIGNAL       1     15
FT   PROPEP      16     22       Activation peptide.
FT   CHAIN        23    146       Phospholipase A2, major isoenzyme.
FT   ACT_SITE     70     70
FT   ACT_SITE    121    121
FT   METAL         50     50       Calcium (via carbonyl oxygen).
FT   METAL         52     52       Calcium (via carbonyl oxygen).
FT   METAL         54     54       Calcium (via carbonyl oxygen).
FT   METAL         71     71       Calcium.
FT   MOD_RES       16     16       Pyrrolidone carboxylic acid.
FT   LIPID         78     78       N6-palmitoyl lysine.
FT   DISULFID      33     99
FT   DISULFID      49    146
FT   DISULFID      51     67
FT   DISULFID      66    127
FT   DISULFID      73    120
FT   DISULFID      83    113
FT   DISULFID     106    118
FT   HELIX         24     34
FT   TURN          36     37
FT   HELIX         40     44
FT   STRAND        47     47
FT   TURN          48     50
FT   STRAND        51     51
FT   HELIX         62     79
FT   HELIX         81     83
FT   TURN          84     85
FT   HELIX         90     92
FT   STRAND        97    100
FT   TURN         101    102
FT   STRAND       103    106
FT   TURN         108    109
FT   HELIX        112    130
FT   HELIX        135    137
FT   STRAND       138    138

Alignment Scores:
Pred. No.:              4.73e-14      Length:          146
Score:                  205.00        Matches:         44
Percent Similarity:     53.9%         Conservative:    11
Best Local Similarity:  43.1%         Mismatches:      18
Query Match:            16.4%         Indels:          29
DB:                     1             Gaps:            3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_PIG (1-146)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
                  ||||||      |||   |||:::                              |||
Db         19 IleSerSerArgAlaLeuTrpGlnPheArgSerMetIleLysCysAlaIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                  ||||||   :::                    :::      :::||| |||   |||
Db         39 HisProLeuMetAspPheAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                  |||:::                                  |||   |||||||||::
Db         58 rProValAsp----------------------------GluLeuAspArgCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
                  :|||||||||||||||||      ||||||   |||||||||||||||||||:::||
Db         68 uThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSerCysLysPheLeuValAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
                  |   |||||||||   :::||||||||||||||||||   :::   |||||||||:::||
Db         88 pAsnProTyrThrGluSerTyrSerTyrSerCysSerAsnThrGluIleThrCysAsnSe 108

Qy        359 TAGG 362
                  |:::
Db        108 rLys 109

RESULT 9
PA22_PIG
ID   PA22_PIG       STANDARD;      PRT;   124 AA.
AC   P04416;
```

```
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2, minor isoenzyme (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC   Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   PRELIMINARY PROTEIN SEQUENCE.
RX   MEDLINE=80088382; PubMed=518908; DOI=10.1016/0005-2795(79)90153-3;
RA   Puijk W.C., Verheij H.M., Wietzes P., de Haas G.H.;
RT   "The amino acid sequence of the phospholipase A2 isoenzyme from
RT   porcine pancreas.";
RL   Biochim. Biophys. Acta 580:411-415(1979).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: This isozyme constitutes about 5% of the pancreatic
CC       phospholipase.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   PIR; A00735; PSPGA2.
DR   HSSP; P00592; 1L8S.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW   Metal-binding.
FT   ACT_SITE     48     48
FT   ACT_SITE     99     99
FT   METAL        28     28       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        30     30       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        32     32       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        49     49       Calcium (By similarity).
FT   DISULFID     11     77
FT   DISULFID     27    124
FT   DISULFID     29     45
FT   DISULFID     44    105
FT   DISULFID     51     98
FT   DISULFID     61     91
FT   DISULFID     84     96
SQ   SEQUENCE   124 AA;  13969 MW;  37E6B121D66FFB2F CRC64;

Alignment Scores:
Pred. No.:              1.01e-13      Length:         124
Score:                  202.00        Matches:        33
Percent Similarity:     85.1%         Conservative:   7
Best Local Similarity:  70.2%         Mismatches:     7
Query Match:            16.1%         Indels:         0
DB:                     1             Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA22_PIG (1-124)

Qy       222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
                 |||   |||||||||:::||||||||||||||||||      ||||||   |||||||||
Db        41 LeuAspArgCysCysGluThrHisAspAsnCysTyrArgAspAlaLysAsnLeuAspSer 60

Qy       282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
                 |||||||||||:::|||   |||||||||::::::||||||||||||||||||   :::
Db        61 CysLysPheLeuValAspAsnProTyrThrAsnSerTyrSerTyrSerCysSerAsnThr 80

Qy       342 GCAATCACCTGTAGCAGTAGG 362
                   |||||||||:::|||:::
Db        81 GluIleThrCysAsnSerLys 87

RESULT 10
PA21B_HORSE
ID   PA21B_HORSE    STANDARD;      PRT;   132 AA.
AC   P00594;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2) (Fragment).
GN   Name=PLA2G1B;
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   PROTEIN SEQUENCE OF 1-7.
RX   MEDLINE=77134902; PubMed=849461; DOI=10.1016/0005-2795(77)90062-9;
RA   Evenberg A., Meyer H., Verheij H.M., de Haas G.H.;
RT   "Isolation and properties of prophospholipase A2 and phospholipase A2
RT   from horse pancreas and horse pancreatic juice.";
RL   Biochim. Biophys. Acta 491:265-274(1977).
RN   [2]
RP   PROTEIN SEQUENCE OF 8-132, AND ACTIVE SITE.
RX   MEDLINE=77118587; PubMed=838712;
RA   Evenberg A., Meyer H., Gaastra W., Verheij H.M., de Haas G.H.;
RT   "Amino acid sequence of phospholipase A2 from horse pancreas.";
RL   J. Biol. Chem. 252:1189-1196(1977).
RN   [3]
RP   SEQUENCE REVISION TO 131-132.
RX   MEDLINE=83283533; PubMed=6349696; DOI=10.1016/0167-4838(83)90126-7;
RA   Verheij H.M., Westerman J., Sternby B., de Haas G.H.;
RT   "The complete primary structure of phospholipase A2 from human
RT   pancreas.";
RL   Biochim. Biophys. Acta 747:93-99(1983).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   PIR; A90615; PSHOA.
DR   HSSP; P00592; 1HN4.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Direct protein sequencing; Hydrolase; Lipid degradation;
KW   Metal-binding.
FT   PROPEP        1      7       Removed by trypsin.
```

```
FT   CHAIN         8    132       Phospholipase A2.
FT   ACT_SITE     55     55       By similarity.
FT   ACT_SITE    106    106       By similarity.
FT   METAL        35     35       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        37     37       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        39     39       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        56     56       Calcium (By similarity).
FT   DISULFID     18     84       By similarity.
FT   DISULFID     34    131       By similarity.
FT   DISULFID     36     52       By similarity.
FT   DISULFID     51    112       By similarity.
FT   DISULFID     58    105       By similarity.
FT   DISULFID     68     98       By similarity.
FT   DISULFID     91    103       By similarity.
FT   VARIANT       1      2       Missing (in a second form).
FT   NON_TER       1      1
SQ   SEQUENCE   132 AA;  14694 MW;  BC6B809A244052EA CRC64;

Alignment Scores:
Pred. No.:                1.17e-13       Length:        132
Score:                    201.50         Matches:       48
Percent Similarity:       43.5%          Conservative:  19
Best Local Similarity:    31.2%          Mismatches:    31
Query Match:              16.1%          Indels:        56
DB:                       1              Gaps:          6

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_HORSE (1-132)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                  |||
Db          4 IleSerProArgAlaValTrpGlnPheArgSerMetIleGlnCysThrIleProAsnSer 23

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
               |||:::|||                  :::      :::||| |||   |||
Db         24 LysProTyrLeuGluPheAsnAspTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 43

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
              |||:::                                  |||      ||||||||
Db         43 rProValAsp-----------------------------GluLeuAspAlaCysCysGl 53

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              |   |||||||||||||||   |||||||||:::|||   ||||||:::||||||:::||
Db         53 nValHisAspAsnCysTyrThrGlnAlaLysGluLeuSerSerCysArgPheLeuValAs 73

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   |||||||||   :::|||   :::|||||||||||||:::   :::||||||||||
Db         73 pAsnProTyrThrGluSerTyrLysPheSerCysSerGlyThrGluValThrCysSerAs 93

Qy        359 TAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAA 418
              :::                                                      ||
Db         93 pLys-----------------------------------------------------As 95

Qy        419 TAATAGT------------------AACAACAGCCATGATTTAGTGTTAATTTTCTTGGT 460
              ||||:::                  |||        :::     :::   |||
Db         95 nAsnAlaCysGluAlaPheIleCysAsnCysAspArgAsnAlaAlaIleCysPhe----- 113

Qy        461 TCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATG 500
                          :::||||||:::|||      :::
Db        114 ----SerLysAlaProTyrAsnProGluAsnLysAsnLeu 125

RESULT 11
PA21B_CAVPO
ID   PA21B_CAVPO    STANDARD;      PRT;   146 AA.
AC   P43434;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (Group IB phospholipase A2).
GN   Name=PLA2G1B;
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC   Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=93345504; PubMed=8344290;
RA   Ying Z., Tojo H., Nonaka Y., Okamoto M.;
RT   "Cloning and expression of phospholipase A2 from guinea pig gastric
RT   mucosa, its induction by carbachol and secretion in vivo.";
RL   Eur. J. Biochem. 215:91-97(1993).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; D00740; BAA00640.1; -; mRNA.
DR   PIR; S34049; S34049.
DR   HSSP; P00592; 1L8S.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding; Signal.
FT   SIGNAL        1     15       By similarity.
FT   PROPEP       16     22       Activation peptide.
FT   CHAIN        23    146       Phospholipase A2.
FT   ACT_SITE     70     70       By similarity.
FT   ACT_SITE    121    121       By similarity.
FT   METAL         50     50       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL         52     52       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL         54     54       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL         71     71       Calcium (By similarity).
FT   DISULFID      33     99       By similarity.
FT   DISULFID      49    146       By similarity.
FT   DISULFID      51     67       By similarity.
FT   DISULFID      66    127       By similarity.
FT   DISULFID      73    120       By similarity.
FT   DISULFID      83    113       By similarity.
FT   DISULFID     106    118       By similarity.
SQ   SEQUENCE   146 AA;  16187 MW;  760DE6CD11E20E07 CRC64;

Alignment Scores:
Pred. No.:                4.17e-12       Length:        146
Score:                    188.00         Matches:       41
Percent Similarity:       53.2%          Conservative:  17
Best Local Similarity:    37.6%          Mismatches:    22
Query Match:              15.0%          Indels:        29
DB:                       1              Gaps:          3

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_CAVPO (1-146)

Qy         92 ATTTTTTCAAGCCATAGGATCATTTCTCCTGAAGCATCTTGGCGA--------------136
              :::   ::::::      ||||||:::|||   |||   |||:::
Db         12 ValSerAlaAlaAlaHisIleIleThrProArgAlaLeuTrpGlnPheArgAspMetIle 31
```

```
Qy        137 ------------------AGTCATCCCCAC--------------------CTGTTCCTG 157
                                  |||   |||:::                        |||
Db         32 LysCysAlaIleProGlySerArgProTyrSerGluTyrAsnAsnTyrGlyCysPheCys 51

Qy        158 AGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCT 217
               :::||| |||   |||   |||:::
Db         52 GlyLeuGly-GlySerGlyThrProValAsp---------------------------- 61

Qy        218 CCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGA 277
                  |||   |||||||||:::   ||||||   ||||||   |||||||||   |||::
Db         62 -GluLeuAspArgCysCysGluIleHisAspAlaCysTyrThrGlnAlaLysHisLeuGl 81

Qy        278 CAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGG 337
                :|||||||||   ::::::|||   |||||||||::::::||||||:::||||||||||
Db         81 uSerCysLysSerValIleAspAsnProTyrThrAsnSerTyrSerPheSerCysSerGl 101

Qy        338 CTCGGCAATCACCTGTAGCAGTAGG 362
               |:::   |||   |||||||||:::
Db        101 yThrAsnIleIleCysSerSerLys 109
```

```
RESULT 12
PA2I_LATSE
ID   PA2I_LATSE     STANDARD;      PRT;   152 AA.
AC   Q8JFB2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 GL16-1 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase) (pkP5).
OS   Laticauda semifasciata (Broad-banded blue sea snake) (Erabu sea
OS   snake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Laticaudinae; Laticauda.
OX   NCBI_TaxID=8631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pancreas;
RX   MEDLINE=22114984; PubMed=12119117; DOI=10.1016/S0378-1119(02)00682-0;
RA   Fujimi T.J., Kariya Y., Tsuchiya T., Tamiya T.;
RT   "Nucleotide sequence of phospholipase A2 gene expressed in snake
RT   pancreas reveals the molecular evolution of toxic phospholipase A2
RT   genes.";
RL   Gene 292:225-231(2002).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides (By similarity).
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AB078348; BAC03247.1; -; mRNA.
DR   EMBL; AB078346; BAC03245.1; -; Genomic_DNA.
DR   HSSP; P00593; 4BP2.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding;
KW   Multigene family; Signal.
FT   SIGNAL        1     21       Potential.
FT   PROPEP       22     27       By similarity.
FT   CHAIN        28    152       Phospholipase A2 GL16-1.
FT   ACT_SITE     75     75       By similarity.
FT   ACT_SITE    126    126       By similarity.
FT   METAL        55     55       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        57     57       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        59     59       Calcium (via carbonyl oxygen) (By
FT                                similarity).
FT   METAL        76     76       Calcium (By similarity).
FT   DISULFID     38    104       By similarity.
FT   DISULFID     54    151       By similarity.
FT   DISULFID     56     72       By similarity.
FT   DISULFID     71    132       By similarity.
FT   DISULFID     78    125       By similarity.
FT   DISULFID     88    118       By similarity.
FT   DISULFID    111    123       By similarity.
SQ   SEQUENCE   152 AA;  16430 MW;  B86D54DC4B77732D CRC64;
```

```
Alignment Scores:
Pred. No.:                1.68e-10        Length:         152
Score:                    174.00          Matches:        29
Percent Similarity:       75.0%           Conservative:   4
Best Local Similarity:    65.9%           Mismatches:     11
Query Match:              13.9%           Indels:         0
DB:                       1               Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA2I_LATSE (1-152)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
                  |||   |||||||||||||||||||||||||||||||||   ||||||||||||      :::
Db         68 LeuAspArgCysCysGlnThrHisAspAsnCysTyrSerGlnAlaLysLysHisProAla 87

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
                  ||||||      ||||||   ||||||         |||||||||:::|||||||||
Db         88 CysLysSerProLeuAspSerProTyrIleLysIleTyrSerTyrThrCysSerGlyGly 107

Qy        342 GCAATCACCTGT 353
                :::::|||||
Db        108 SerLeuThrCys 111
```

```
RESULT 13
PA2H_LATSE
ID   PA2H_LATSE     STANDARD;      PRT;   152 AA.
AC   Q8JFG2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 pkP2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
DE   acylhydrolase).
OS   Laticauda semifasciata (Broad-banded blue sea snake) (Erabu sea
OS   snake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Laticaudinae; Laticauda.
OX   NCBI_TaxID=8631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pancreas;
RX   MEDLINE=22114984; PubMed=12119117; DOI=10.1016/S0378-1119(02)00682-0;
RA   Fujimi T.J., Kariya Y., Tsuchiya T., Tamiya T.;
RT   "Nucleotide sequence of phospholipase A2 gene expressed in snake
RT   pancreas reveals the molecular evolution of toxic phospholipase A2
RT   genes.";
RL   Gene 292:225-231(2002).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides (By similarity).
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
```

```
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AB078347; BAC03246.1; -; mRNA.
DR   HSSP; P00593; 4BP2.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding;
KW   Multigene family; Signal.
FT   SIGNAL      1     21       Potential.
FT   PROPEP      22    27       By similarity.
FT   CHAIN       28    152      Phospholipase A2 pkP2.
FT   ACT_SITE    75    75       By similarity.
FT   ACT_SITE    126   126      By similarity.
FT   METAL       55    55       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       57    57       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       59    59       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       76    76       Calcium (By similarity).
FT   DISULFID    38    104      By similarity.
FT   DISULFID    54    151      By similarity.
FT   DISULFID    56    72       By similarity.
FT   DISULFID    71    132      By similarity.
FT   DISULFID    78    125      By similarity.
FT   DISULFID    88    118      By similarity.
FT   DISULFID    111   123      By similarity.
SQ   SEQUENCE   152 AA;  16374 MW;  F763BAD02C11BFEB CRC64;

Alignment Scores:
Pred. No.:              4.83e-10      Length:         152
Score:                  170.00        Matches:        28
Percent Similarity:     76.2%         Conservative:   4
Best Local Similarity:  66.7%         Mismatches:     10
Query Match:            13.6%         Indels:         0
DB:                     1             Gaps:           0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA2H_LATSE (1-152)

Qy        228 AGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAA 287
              ||||||||||||||||||||||||||||  ||||||||||||      :::||||||
Db         70 ArgCysCysGlnThrHisAspAsnCysTyrSerGlnAlaLysLysHisProAlaCysLys 89

Qy        288 TTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATC 347
                  ||||||   ||||||          |||||||||:::|||||||||   ::::::
Db         90 SerProLeuAspSerProTyrIleLysIleTyrSerTyrThrCysSerGlyGlySerLeu 109

Qy        348 ACCTGT 353
              ||||||
Db        110 ThrCys 111

RESULT 14
PA21B_OPHHA
ID   PA21B_OPHHA    STANDARD;      PRT;   151 AA.
AC   P80966; Q9DF32;
DT   01-NOV-1997 (Rel. 35, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2, acidic 1 precursor (EC 3.1.1.4) (Phosphatidylcholine
DE   2-acylhydrolase) (APLA2-1) (OHV A-PLA2) (OHV-APLA2).
OS   Ophiophagus hannah (King cobra) (Naja hannah).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Elapinae; Ophiophagus.
OX   NCBI_TaxID=8665;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Venom gland;
RA   Wang Q.Y., Shu Y.Y.;
RT   "Cloning and characterization of cDNAs encoding two acidic isoforms of
RT   phospholipase A2 in Guangxi King Cobra (Ophiophagus hannah).";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   PROTEIN SEQUENCE OF 28-151.
RC   TISSUE=Venom;
RX   MEDLINE=97181016; PubMed=9028866; DOI=10.1006/abbi.1996.9814;
RA   Huang M.Z., Gopalakrishnakone P., Chung M.C.M., Kini R.M.;
RT   "Complete amino acid sequence of an acidic, cardiotoxic phospholipase
RT   A2 from the venom of Ophiophagus hannah (King Cobra): a novel cobra
RT   venom enzyme with 'pancreatic loop'.";
RL   Arch. Biochem. Biophys. 338:150-156(1997).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides. Exhibits cardiotoxicity,
CC       myotoxicity, antiplatelet activity, and edema-inducing activity.
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AF302908; AAG23964.1; -; mRNA.
DR   PDB; 1GP7; X-ray; A/B/C=1-151.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   3D-structure; Calcium; Cardiotoxin; Direct protein sequencing;
KW   Hydrolase; Lipid degradation; Metal-binding; Multigene family; Signal;
KW   Toxin.
FT   SIGNAL      1     21       Potential.
FT   PROPEP      22    27
FT   CHAIN       28    151      Phospholipase A2, acidic 1.
FT   ACT_SITE    75    75       By similarity.
FT   ACT_SITE    126   126      By similarity.
FT   METAL       55    55       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       57    57       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       59    59       Calcium (via carbonyl oxygen) (By
FT                              similarity).
FT   METAL       76    76       Calcium (By similarity).
FT   DISULFID    38    104      By similarity.
FT   DISULFID    54    151      By similarity.
FT   DISULFID    56    72       By similarity.
FT   DISULFID    71    132      By similarity.
FT   DISULFID    78    125      By similarity.
FT   DISULFID    88    118      By similarity.
FT   DISULFID    111   123      By similarity.
SQ   SEQUENCE   151 AA;  16444 MW;  F1E021B886BC4BEB CRC64;

Alignment Scores:
Pred. No.:              5.16e-09      Length:         151
Score:                  161.00        Matches:        25
```

```
Percent Similarity:     67.4%          Conservative:  6
Best Local Similarity:  54.3%          Mismatches:    15
Query Match:            12.8%          Indels:        0
DB:                     1              Gaps:          0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA21B_OPHHA (1-151)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||   ||||||||||||||   ||||||:::||||||   :::
Db         68 LeuAspArgCysCysGlnValHisAspAsnCysTyrThrGlnAlaGlnLysLeuProAla 87

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              |||      ::::::|||   ||||||         |||||||||   ||||||
Db         88 CysSerSerIleMetAspSerProTyrValLysIleTyrSerTyrAspCysSerGluArg 107

Qy        342 GCAATCACCTGTAGCAGT 359
                 :::||||||   :::
Db        108 ThrValThrCysLysAla 113

RESULT 15
PA217_AUSSU
ID   PA217_AUSSU    STANDARD;      PRT;   152 AA.
AC   Q9PUG7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Phospholipase A2 isozyme S17-58 precursor (EC 3.1.1.4)
DE   (Phosphatidylcholine 2-acylhydrolase) (ASPLA17).
OS   Austrelaps superbus (Australian copperhead).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Acanthophiinae; Austrelaps.
OX   NCBI_TaxID=29156;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Venom gland;
RX   MEDLINE=20166950; PubMed=10700385; DOI=10.1006/abbi.1999.1672;
RA   Singh S.B., Armugam A., Kini R.M., Jeyaseelan K.;
RT   "Phospholipase A(2) with platelet aggregation inhibitor activity from
RT   Austrelaps superbus venom: protein purification and cDNA cloning.";
RL   Arch. Biochem. Biophys. 375:289-303(2000).
CC   -!- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the 2-
CC       acyl groups in 3-sn-phosphoglycerides. Inhibits collagen-induced
CC       platelet aggregation (By similarity).
CC   -!- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC       acylglycerophosphocholine + a carboxylate.
CC   -!- COFACTOR: Calcium (Probable).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Belongs to the phospholipase A2 family. Group I
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AF184143; AAD56410.1; -; mRNA.
DR   HSSP; P00593; 4BP2.
DR   InterPro; IPR001211; PhospholipaseA2.
DR   PANTHER; PTHR11716; PhospholipaseA2; 1.
DR   Pfam; PF00068; Phospholip_A2_1; 1.
DR   PRINTS; PR00389; PHPHLIPASEA2.
DR   ProDom; PD000303; PhospholipaseA2; 1.
DR   SMART; SM00085; PA2c; 1.
DR   PROSITE; PS00119; PA2_ASP; 1.
DR   PROSITE; PS00118; PA2_HIS; 1.
KW   Calcium; Hydrolase; Lipid degradation; Metal-binding;
KW   Multigene family; Signal; Toxin.
FT   SIGNAL        1     19       Potential.
FT   PROPEP       20     27       Potential.
FT   CHAIN        28    152       Phospholipase A2 isozyme S17-58.
FT   ACT_SITE     75     75       By similarity.
FT   ACT_SITE    126    126       By similarity.
FT   METAL         55     55       Calcium (via carbonyl oxygen) (By
FT                                 similarity).
FT   METAL         57     57       Calcium (via carbonyl oxygen) (By
FT                                 similarity).
FT   METAL         59     59       Calcium (via carbonyl oxygen) (By
FT                                 similarity).
FT   METAL         76     76       Calcium (By similarity).
FT   DISULFID      38    104       By similarity.
FT   DISULFID      54    151       By similarity.
FT   DISULFID      56     72       By similarity.
FT   DISULFID      71    132       By similarity.
FT   DISULFID      78    125       By similarity.
FT   DISULFID      88    118       By similarity.
FT   DISULFID     111    123       By similarity.
SQ   SEQUENCE   152 AA;  16641 MW;  0F2BD5BDDCF04FCB CRC64;

Alignment Scores:
Pred. No.:              5.17e-09       Length:        152
Score:                  161.00         Matches:       27
Percent Similarity:     71.1%          Conservative:  5
Best Local Similarity:  60.0%          Mismatches:    13
Query Match:            12.8%          Indels:        0
DB:                     1              Gaps:          0

US-10-607-806-1-G7328_COPY_7000_7700 (1-701) x PA217_AUSSU (1-152)

Qy        222 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 281
              |||   ||||||||||||||||||||||||||||   :::|||   ||||||   :::
Db         68 LeuAspArgCysCysGlnThrHisAspAsnCysTyrAlaGluAlaGlyLysLeuProAla 87

Qy        282 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCG 341
              ||||||   :::|||      ||||||      ||||||||||||||||
Db         88 CysLysAlaMetLeuSerGluProTyrAsnAspThrTyrSerTyrSerCysIleGluArg 107

Qy        342 GCAATCACCTGTAGC 356
                 :::||||||:::
Db        108 GlnLeuThrCysAsn 112


Search completed: February  8, 2006, 11:14:13
Job time : 164.625 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:43:51 ; Search time 14.7309 Seconds
(without alignments)
2613.961 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 3524
Sequence: 1 ctatctcagctgtccctccc.....ckctacataaatatctatta 2001

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091821_26876/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_92_@runat_08022006_091821_26876 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		416	11.8	673	4	F40201	artifact-warning s
2	c	391.5	11.1	673	4	F40201	artifact-warning s
3	c	384.5	10.9	613	4	C40201	artifact-warning s
4		366	10.4	613	4	C40201	artifact-warning s
5		332.5	9.4	627	4	A40201	artifact-warning s
6	c	322	9.1	627	4	A40201	artifact-warning s
7	c	267	7.6	597	4	E40201	artifact-warning s
8		264	7.5	196	2	I38022	hypothetical prote
9		259.5	7.4	597	4	E40201	artifact-warning s
10		251	7.1	148	1	PSHU	phospholipase A2 (
11	c	248.5	7.1	301	4	B40201	artifact-warning s
12		239.5	6.8	301	4	B40201	artifact-warning s
13		236.5	6.7	579	4	D40201	artifact-warning s
14	c	232.5	6.6	100	2	A46010	X-linked retinopat
15		227	6.4	418	2	S41044	chromosomal protei
16	c	223.5	6.3	579	4	D40201	artifact-warning s
17		217	6.2	146	1	PSDG	phospholipase A2 (
18		216	6.1	146	1	PSRT	phospholipase A2 (
19	c	214.5	6.1	499	2	S65657	alpha-1C-adrenergi
20		209	5.9	145	1	PSBOA	phospholipase A2 (
21		208	5.9	144	2	JN0480	phospholipase A2 (
22	c	206.5	5.9	418	2	S41044	chromosomal protei
23	c	206	5.8	196	2	I38022	hypothetical prote
24		202	5.7	124	1	PSPGA2	phospholipase A2 (
25		201	5.7	146	1	PSPGA	phospholipase A2 (
26		191.5	5.4	132	1	PSHOA	phospholipase A2 (
27		177	5.0	146	1	S34049	phospholipase A2 (
28	c	172	4.9	46	2	I54375	gene NF2 protein -
29	c	162	4.6	39	2	I54374	gene NF2 protein -
30	c	160	4.5	369	2	A53959	thromboxane A-2 re
31		157	4.5	125	2	S65624	phospholipase A(2)
32	c	156	4.4	53	2	A42442	integrin beta-1 ch
33		155.5	4.4	119	2	S00227	phospholipase A2 (
34	c	155.5	4.4	407	2	T02670	probable thromboxa
35	c	154.5	4.4	79	2	A56194	thromboxane A-2 re
36		149.5	4.2	119	1	PSNJ3K	phospholipase A2 (
37	c	147	4.2	841	1	I78885	serine/threonine-s
38		144.5	4.1	146	1	PSNJAF	phospholipase A2 (
39		144.5	4.1	499	2	S65657	alpha-1C-adrenergi
40		143.5	4.1	119	1	PSNJ2K	phospholipase A2 (
41	c	143.5	4.1	124	2	A47582	B-cell growth fact
42		142.5	4.0	100	2	A46010	X-linked retinopat
43		142	4.0	133	1	PSOXG	phospholipase A2 (
44		140.5	4.0	119	1	PSRIA	phospholipase A2 (
45		138.5	3.9	118	1	PSKFT3	phospholipase A2 (

ALIGNMENTS

RESULT 1
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of
.

Alignment Scores:
Pred. No.: 2.89e-35 Length: 673
Score: 416.00 Matches: 131
Percent Similarity: 41.8% Conservative: 17
Best Local Similarity: 37.0% Mismatches: 85
Query Match: 11.8% Indels: 122
DB: 4 Gaps: 9

US-10-607-806-1_COPY_7200_9200 (1-2001) x F40201 (1-673)

```
Qy      332 AAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGG 391
            ||||||        ||||||              |||         |||   |||
Db      328 LysLysLysAsnProGlnLys--------------Gln*********PheCysPheTrp 342

Qy      392 TGACAGTTTGGGTTTTTTTTTGTTGTTGTTTA-------GAGACAGGGTCTTGCTCTGTC 444
```

```
Db        343 GlyPhePhePhePhePhePheCysPheValPheValPheGluIleGlyProCysSerAsp 362

Qy        445 ACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTC 504

Db        363 ThrProAlaGlyValGlnTrp-GlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySe 382

Qy        505 AAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGC 564

Db        382 rSerAspSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSe 402

Qy        565 CTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGT 624

Db        402 rTrpLeuIleTyrVal--PheLeuIleGluThrGlnPheHisHisValAspGlnAlaGly 421

Qy        625 CTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACA 684

Db        422 LeuLysLeuLeuThrSerSerAspLeuProSer-TrpSerProLysValLeuGly***Gl 441

Qy        685 GGGGTGAGCCACC----------ATGTGCGGCTACTTATTTCTTTA 720

Db        441 nAla***AlaThrThrProSer*********PheValPheGlyGlyPhePhePhePhPh 461

Qy        721 CATTCCA---------- 727

Db        461 ePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArgLeuGluCysSe 481

Qy        728 ----TCTTTCCAATAGAATGTAAGATCCACA---------GAACAGGGATTACTGCCTAT 774

Db        481 rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLe 501

Qy        775 T----------TTCTTCCTTTC 786

Db        501 uSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyrPh 521

Qy        787 TTTTTTGAGACAGAGTCTCACT----------TCATCACCTCAA-- 820

Db        521 e******ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVa 541

Qy        821 ----------CCTCCG-------- 826

Db        541 lIleCysProProGlyLeuProLysCysTrpAspAspArgArgGluProProHisProAl 561

Qy        826 ---------- 826

Db        561 a*********LeuPheLeuGlyValPhePhePhePhePheLeuLeuCysPheCysPhe** 581

Qy        827 ----------TTCAGCTCACTGCA 840

Db        581 *AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuGl 601

Qy        841 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 900

Db        601 nProGlnProProGlyPheLysArgPheSerCysLeuCysLeuProCysSerTrpAspTy 621

Qy        901 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 960

Db        621 rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP 641

Qy        961 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1020

Db        641 roCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSer 660

Qy       1021 CAAAGTGCTGGAATTATAGGCGTGAGTCAC 1050

Db        661 GlnSerAlaGlyMetThrGlyValSerHis 670

RESULT 2
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                1.15e-32      Length:          673
Score:                    391.50        Matches:         122
Percent Similarity:       57.9%         Conservative:    21
Best Local Similarity:    49.4%         Mismatches:      65
Query Match:              11.1%         Indels:          41
DB:                       4             Gaps:            5

US-10-607-806-1_COPY_7200_9200 (1-2001) x F40201 (1-673)

Qy       1046 CTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAGG 987

Db          6 LeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHisLeuArgSerGl 25

Qy        986 AGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAAT 927

Db         25 yVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLysIleHis-LysL 45

Qy        926 TAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAGA 867

Db         45 euAlaArgAsnGlyGlyProCysLeu***SerGlnLeuHisGlyArgGlnArgGlnGluA 65

Qy        866 ATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG----------CTGAACG 825

Db         65 snArgLeuAsnProGlyGly***GlyCysSerGluProLysLeuAlaThrAlaLeuGlnP 85

Qy        824 GAGGTTG-AGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGT 766

Db         85 roGlyCysGln------SerLysGlyLeuSerGlnLysGlnLysGlnSerLysLysLys- 102

Qy        765 AATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTA- 707

Db        103 ----------LysLysLysThrP 107

Qy        706 ----------GCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTG 667

Db        107 roLysAsnLys*********AlaGlyCysGlyGlySerArgLeuSerSerGlnHisPheG 127

Qy        666 GGATGCCAAGGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTA 607

Db        127 lyArgPro-GlyGlyGlnIleThr***GlyGlnGluPheGluThrSerLeuIleAsnMet 146

Qy        606 GTAAGACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTC 547

Db        147 ValLysLeuCysLeuTyr***Lys--TyrIleAsn***ProGlyMetAlaAlaHisAlaC 166

Qy        546 GTAGTCCCAGCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGG 487

Db        166 ysAsnProSerTyrThrGlyAspArgGlyArgArgIleAla***ThrGlnGluValGluA 186

Qy        486 CTGCAGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGT 427

Db        186 laAlaValSerGlnAsn-LeuProLeuHisSerSerArgGlyValArgAlaArgAlaTyr 205

Qy        426 CTCTA-------AACAACAACAAAAAAAAACCCAAACTGTCACCACTTACAAGCTGTGTG 374
```

```
Db        206 LeuLysAsnLysAsnLysAlaLysLysLysLysLysProProLysThrLys****** 225

Qy        373 ATCTTGGGC 365
              |||||
Db        226 ***LeuGly 228
```

RESULT 3
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	6.29e-32	Length:	613
Score:	384.50	Matches:	117
Percent Similarity:	52.0%	Conservative:	24
Best Local Similarity:	43.2%	Mismatches:	79
Query Match:	10.9%	Indels:	51
DB:	4	Gaps:	4

US-10-607-806-1_COPY_7200_9200 (1-2001) x C40201 (1-613)

```
Qy       1081 AAAGAAAATAGACAGTAATCGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTT 1022
              |||:::   :::          |||      |||   |||||||||      |||||||||
Db         97 LysLysLysLys*********AlaGlyCysSerGlySerArgLeu***SerGlnHisPhe 116

Qy       1021 GGGAGACTGAGGCAGGCAGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACAT 962
              ||||||   |||      ||||||  |||   ||||||   ||||||||||||   ||||
Db        117 GlyArgGlnArgTrpValAspHis--GluAlaArgSerSerArgProAlaTrpProThrT 136

Qy        961 GGTAAAACCCCATCTCTGCTAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTT 902
              ||   ||||||||||||   ||||||||||||||          ||||||
Db        136 rpGlnAsnProIleSerThrLysAsnThrLysAsn***LeuGlyMetVal-ArgAlaPro 155

Qy        901 GTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGG 842
              |||:::|||   ||||||:::|||   |||||||||      :::|||||||||
Db        156 ValValProArgThrGlnLysAlaGluAlaGlyGluTrpHisLysProGlyArgTrpSer 175

Qy        841 TTGCAGTGAGCTGAA--------------CGGAGGTTGAGGTGATGAAGTGAGACTCTG 797
              ||||||   ||||||                  ||| ||||||   ||||||||||
Db        176 LeuGln***AlaGluIleSerProLeuHisSerSerLeu-GlyAspArgValArgLeuCy 195

Qy        796 TCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTA 737
              ||||||||||::::::||||||
Db        195 sLeuLysLysLysLysLys---------------------------------------- 202

Qy        736 TTGGAAAGATGGAATGTAAAGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCC 677
                          :::                  |||||||||   ||||||||
Db        203 ------------------Lys*********ArgGlyAlaValAlaHisAlaCysAsnPr 216

Qy        676 CAGCACTTTGGGATGCCAAGGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCT 617
              |||||||||||||   :::||| |||   ::: :::::||||||||   |||:::|||
Db        216 oSerThrLeuGlyGlyLysGly-GlyTrpIle-MetArgProGlyValArgAspGlnPro 235

Qy        616 GGCCAACGTAGTAAGACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTG 557
              ||||||   ||||||   |||   ||||||   ||||||   ||||||
```

```
Db        236 GlyGlnHisGlyLysThrProPheLeu-LeuLysIleGlnLysIleSerTrpAlaTrpCy 255

Qy        556 GTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAG 497
                  |||      |||   :::      |||||||||      :::   ::::::   |
Db        255 s-GlyArgLeu***SerHisValArgArgArgLeuArgGlnGluAsnGlyIleAsnProG 275

Qy        496 GAGGTTGAGGCTGCAGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGC 437
              |||||      ||||||:::            ||||||      ||||||||||||||||
Db        275 lyGlyGlyAlaCysSerGluProArgSerArgHisCysThrProAlaTrpValThrGlu* 295

Qy        436 AAGACCCTGTCTCTAAACAACAACAAAAAAAA---------------------------- 405
                |||   |||||||||:::::::::||||||
Db        295 **AspSerValSerLysLysLysLysLys*********PhePhePhePhePhePheP 315

Qy        404 ----------------------ACCCAAACTGTCACCACTTACAAGCTGTGTGATCTTG 368
                                       ||||||          ::::::      |||||||
Db        315 heGluThrGluSerHisSerValThrGlnAlaGlyValGlnTrpArg------AspLeuG 333

Qy        367 GGCAAATAGCTCAGCCT 351
              ||   :::      |||
Db        333 lySerLeuGlnAlaPro 338
```

RESULT 4
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentia
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of
.

Alignment Scores:

Pred. No.:	5.77e-30	Length:	613
Score:	366.00	Matches:	134
Percent Similarity:	40.8%	Conservative:	17
Best Local Similarity:	36.2%	Mismatches:	129
Query Match:	10.4%	Indels:	90
DB:	4	Gaps:	6

US-10-607-806-1_COPY_7200_9200 (1-2001) x C40201 (1-613)

```
Qy        199 CTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTG 258
              :::   |||||||||      :::             |||                |||
Db        226 IleMetArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeu 245

Qy        259 GTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTACAATTATCC 318
                 :::      :::|||:::                          |||:::
Db        246 LysIleGlnLysIleSerTrpAlaTrpCysGlyArgLeu***SerHisValArgArgArg 265

Qy        319 TCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACAC 378
                    |||   ||||||      :::   |||          |||
Db        266 Leu------ArgGlnGluAsnGlyIleAsnProGlyGlyGlyAlaCysSerGluProArg 283

Qy        379 AGCTTGTAAGTGGTGACAGTTTGGGTT------------------------------- 405
              |||                |||||
Db        284 SerArgHisCysThrProAlaTrpValThrGlu***AspSerValSerLysLysLysLys 303

Qy        406 -----------------TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACC 447
                              ||||||              ||||||   |||   |||||||||
```

```
Db        304 LysLys*********PhePhePhePhePhePheGluThrGluSerHisSerValThr 323

Qy        448 CAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAG 507
              ||||||      ||||||   :::      |||||||||      ||||||
Db        324 GlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyPheMet 343

Qy        508 GGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTG 567
                 :::   |||||||||      :::|||||||||   |||    |||||||||
Db        344 ProPheSerCysLeuSerLeuLeuArgThrTrpAspTyrArgArgPro-HisHisAlaGl 363

Qy        568 GCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTT 627
              ||||||      ||| |||   |||:::   ||||||      |||||||||   |||
Db        363 nLeuIlePheCysIle-PheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerA 383

Qy        628 AAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGG 687
              |||||||||        |||      |||      ||||||||||||||||||||
Db        383 rgThrProGlyLeu--MetIleHisProProLeuProProLysValLeuGlyLeuGlnAl 402

Qy        688 GTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 747
                 ||||||         :::   |||
Db        402 a***AlaThrAlaProArgPhePhePhe------------------------------ 411

Qy        748 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 807
                                       ||||||        ||||||||||||||||
Db        412 ------------------------PhePhePhe*********LeuArgGlnSerLeuTh 423

Qy        808 TTCATCACCTCAACCTCCGT---------------TCAGCTCACTGCAACCTCTGCCTCC 852
              |   ||||||::: |||                  ||||||||||||   |||   ||||
Db        423 rLeuSerProArg-LeuGluCysAsnGlyGluIleSerAlaHisCysLysLeuHisLeuP 443

Qy        853 CGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACC 912
              |||||        |||||||||   |||      |||   |||   |||
Db        443 roGlyLeuCysHisSerProAlaSerAlaPheCysValArgGlyThrThrGlyAlaArg- 462

Qy        913 ACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAG 972
              ||||||          ||||||||||||   ||||||||||   ||||||   |||
Db        463 ThrMetProSer***PhePheValPheLeuValGluMetGlyPheCysHisValGlyGln 482

Qy        973 GCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGA 1032
              |||||||||:::||||||   |||      |||      |||   |||||||||||||:
Db        483 AlaGlyLeuGluLeuLeuAlaSer***--SerThrHisLeuCysLeuProLysCysTrpA 502

Qy       1033 ATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATC 1092
              ::|||||||||||||   |||      |||
Db        502 spTyrArgArgGluProLeuHisProAla******------------------------ 513

Qy       1093 CCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCA 1152

Db        513 ------------------------------------------------------------ 513

Qy       1153 CAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAAT 1212
                                  |||||||||||||||                |||   :::
Db        514 --------------***PhePhePhePhePhe-----------------Phe***AspA 523

Qy       1213 AAATACAATCTTGC 1226
                 :::      |||
Db        523 rgValSerLeuCys 527
```

RESULT 5
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	2.06e-26	Length:	627
Score:	332.50	Matches:	114
Percent Similarity:	46.6%	Conservative:	18
Best Local Similarity:	40.3%	Mismatches:	86
Query Match:	9.4%	Indels:	65
DB:	4	Gaps:	4

US-10-607-806-1_COPY_7200_9200 (1-2001) x A40201 (1-627)

```
Qy        397 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 456
              :::::::::||||||            ||||||   |||   ||||||   :::|||
Db        320 LeuPheLeuPhePheSer---------GluThrGluSerArgSerValAlaLysAlaGly 336

Qy        457 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 516
                  ||||||   :::      |||||||||      |||||                  :::
Db        337 ValGlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyPheThrProPheSer 356

Qy        517 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 576
                 ||||||||||||||||||||||||||||   |||    |||:::|||   ||||||
Db        357 CysLeuSerLeuProSerSerTrpAspTyrArgArgPro-HisTyrAlaArgLeuIlePh 376

Qy        577 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 636
                 ||| |||   ||||||   |||              ||||||   |||   :::|||
Db        376 eCysIle-PheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSerProA 396

Qy        637 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 696
                    |||   ||||||   |||||||||||||||||||||
Db        396 spLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGlnAla-------- 412

Qy        697 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 756
                                                          |||:::
Db        413 ------------------------------------------***AspArgAlaArg** 418

Qy        757 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 816
                                       |||      |||   |||:::||||||     |||||
Db        418 *******PhePheIlePheCysPheTyrPhePheProArgArgSerLeuAlaLeuSerPr 438

Qy        817 TCAACCTCCGT---------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 861
              |::: |||                  ||||||||||||   |||   ||||||||||||
Db        438 oArg-LeuGluCysSerGlyAlaIleSerAlaHisCysLysLeuArgLeuProGlySerA 458

Qy        862 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 921
                 |||||||||   ||||||   |||||||||   |||      |||
Db        458 rgHisSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyAlaArg-ThrThrPro 477

Qy        922 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 981
              |||   |||||||||||||||   |||   ||||||:::      :::|||   ||||||
Db        478 Gly***PhePheValPheLeuValGluThrGlyPheHisArgGlySerGlnAspGlyLeu 497

Qy        982 AAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGC 1041
                 ||||||||||||   ||||||   |||   |||||||||||||:::|||||||
Db        498 AspLeuLeuThrSer***--SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgA 517

Qy       1042 GTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCT 1101
              |||||  |||   ||||||           :::||||||:::
Db        517 rgGlu--ThrAlaProGly*********PheLeuPhePheVal----------------- 530

Qy       1102 AGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGA 1161

Db        530 ------------------------------------------------------------ 530

Qy       1162 TATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGA 1210
                       |||:::|||   :::            |||||||||
Db        531 ------------PheIlePhePheArgAspGlyValSerLeuCysArg 542
```

```
RESULT 6
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                2.68e-25      Length:          627
Score:                    322.00        Matches:         104
Percent Similarity:       54.6%         Conservative:    20
Best Local Similarity:    45.8%         Mismatches:      73
Query Match:              9.1%          Indels:          30
DB:                       4             Gaps:            2

US-10-607-806-1_COPY_7200_9200 (1-2001) x A40201 (1-627)

Qy       1047 ACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAG 988
                 :::||||||   |||   ||||||||||||      |||   :::||||||  |||||||
Db          5 SerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyGlyArgIleThr--ArgSerG 24

Qy        987 GAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAA 928
                ||      |||   ||||||   ||||||:::|||||||||||||||||||||||||||||
Db         24 lyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleGlnLysI 44

Qy        927 TTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAG 868
                |||||:::|||          ||||||     |||||||||   ||||||||||||||||
Db         44 leSerArgAla-***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu 63

Qy        867 AATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG---------------CTGAACGGA 823
                |||   :::|||||||||||||      ||||||||                    |||
Db         64 AsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys-ThrPr 83

Qy        822 GGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAAT 763
                        :::   ||||||||||||:::|||
Db         83 oAlaLeuAlaThrGluArgAspSerValSerGluLys---------------------- 95

Qy        762 CCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTAGCCG 703
                                          |||   :::      :::
Db         96 ----------------------------AsnLysAsnLysLys***Lys*********Al 106

Qy        702 CACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGATTGCT 643
                  ||||||||||||||||||||||||||||:::||||||   ||||||  :::
Db        106 aGlyArgGlyLeuThrProValIleProAlaLeuTrpGluAlaLysAlaGlyGly--Ser 125

Qy        642 TGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAAAAAA 583
                  ||||||   :::|||   ||||||      |||:::|||   ||||||   ||
Db        126 ArgGlyGlnGluIleGluThrIleLeuAlaThrThrValLysProArgLeuTyr-***Ly 145

Qy        582 ATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCT 523
                |:::      ||||||   |||           |||   ||||||||||||||||||||
Db        145 sTyrLysLysLeuAlaGlyArgSerAlaGly-AlaCysSerProSerTyrLeuGlyGly* 165

Qy        522 GAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCACCA 463
                  |||:::::::::          :::|||   |||   ||||||:::          |
Db        165 **GlyArgArgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaT 185

Qy        462 CTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAACAACAAAAAAAAAC 403
                  ||   |||   |||       ||||||:::|||   |||              |||   |||
Db        185 hrAlaLeuGlnProTrpArgGlnSerGluThrProSerArgLysLysIleLysThrLysA 205

Qy        402 CCAAA 398
                |||
Db        205 snLys 206

RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of
.

Alignment Scores:
Pred. No.:                1.82e-19      Length:          597
Score:                    267.00        Matches:         135
Percent Similarity:       42.5%         Conservative:    30
Best Local Similarity:    34.8%         Mismatches:      121
Query Match:              7.6%          Indels:          105
DB:                       4             Gaps:            13

US-10-607-806-1_COPY_7200_9200 (1-2001) x E40201 (1-597)

Qy       1061 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1002
                 ||||||          |||   |||:::|||||||||||||||||   :::|||||||||
Db          1 GlyGlnValTrp***LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArg 20

Qy       1001 TC-ACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGC 943
                 |||   |||   |||||||||||||||||   ||||||   ||||||||||||
Db         21 IleThr***GlyTrpGluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr 40

Qy        942 TAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGG 883
                 ||||||  |||||||||   |||||||||          |||||||||||||||:::|
Db         41 ***LysTyr--LysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgG 60

Qy        882 AGGCTTAGGCAG---GAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTG--- 829
                 |||||           ||||||   ||||||||||||   :::   |||   |||:::
Db         60 luAla-GluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGluIleVal 79

Qy        828 ------AACGGAGGTTGAGGTGATGA-AGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAA 776
                       |||   |||:::   |||:::|||   |||||||||:::|||
Db         80 ProSerTyrSerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys****** 99

Qy        775 AATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAG 716
                    :::             |||||||||
Db        100 ***AlaArg-----------CysGlySer---------------------------- 105

Qy        715 AAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGG 656
                                    |||   |||   ||||||   ||||||      |||
Db        106 ----------------------SerCysLeu***SerGlnArgPheGlySerSerArg 117

Qy        655 TA-GGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 597
                  ||||||      |||      |||||||||   |||   ||||||      :::|||
Db        118 ArgGlyGlySerProGluValGlySerLeuArgProAla***ProThrTrpArgAsnPro 137
```

```
Qy        596 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 537
              :::|||||||||           ||||||  ||||||||||||        ||||||
Db        138 IleSerThrLysAsnIleLys---LeuAlaGlyArgGlyGlyAlaCysLeu***SerGln 156

Qy        536 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 477
              ||||||||||||||||||      :::   ::::::
Db        157 LeuLeuGlyArgLeuArgGln---GluArgIleAla***ThrGlnGluAlaLysPheAla 175

Qy        476 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAA------------------ 435
              :::   |||   |||          ||||||   |||
Db        176 ValArgLeuCysHisArgThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys 195

Qy        434 ----------------------------------------GACCCT--------------- 429
                                                     :::|||
Db        196 LysLys*********ProGlyValValAlaHisAlaCysAsnProSerAlaLeuGlyVal 215

Qy        428 -----------------------------GTCTCTAAACAACAACAAAAAAAAACCCAA 399
                                          |||      |||      :::      :::
Db        216 GlnGlyGlyAlaAspHisLeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGlu 235

Qy        398 ACTGTCACCACTTACAAGCTGTGTGAT---CTTGGGCAAATAGCTCAGCCTTCTGAGCCT 342
              |||   :::      |||:::   :::   ||||||  :::|||                |||
Db        236 ThrProSerLeuLeuLysIle***Asn***LeuGlyValValAlaHisAlaCysAsnPro 255

Qy        341 CAGTTTTCTTATCTGTTAAGTGAGGATAATTGTACCTATCCCATAGTGTTGTTCTGAGGA 282
               |||:::
Db        256 -SerTyr----------------------------------------SerGlyGly** 261

Qy        281 TTAAAGGAGACACTGCCCAGAACCAAGAAAATTAACACTAAATCATGGCTGTTGTTACTA 222
              ||||||          ||||||||||||   :::|||          |||      :::
Db        261 *AspArgArgGluLeuLeuGluProArgLysGlnSerLeuArg***AspCysAla--Ile 280

Qy        221 TTATTTCCCCCCGGCCTACTGAGAACCAACTAGAATTCATAGGTCAAG------------ 174
              :::      ||||||          |||      |||      |||
Db        281 ValLeuArgProGlyGlnGlnGluGlnAsnSerValSerLysLysLysLys********* 300

Qy        173 ---------------GAAGGGATAAACCTACTGCTACAGGTGATTGCCGAGCCAGAGCAY 129
                            :::|||:::   |||||||||                |||
Db        301 PhePhePhePhe***AspGlyValLeuLeuLeuLeu--------------ProArgPro 315

Qy        128 GAGTATGAATAGGTGTGGGTGTACGGKKTGTCCAGCAGAAATTTACAGCTGTCCAGCTTC 69
              ||||||:::                |||      |||              |||      |||
Db        316 GluTyrAsp---------------GlyThrIleSerProGlnThrLeuLeuProGlyPhe 330

Qy         68 TTGGCCTGGTCR 57
                 :::|||
Db        331 LysGlnPheSer 334
```

RESULT 8
I38022
hypothetical protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C;Accession: I38022
R;Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A;Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A;Reference number: I38021
A;Accession: I38022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI

Alignment Scores:

Pred. No.:	3.43e-19	Length:	196
Score:	264.00	Matches:	61
Percent Similarity:	68.2%	Conservative:	12
Best Local Similarity:	57.0%	Mismatches:	23
Query Match:	7.5%	Indels:	12
DB:	2	Gaps:	3

US-10-607-806-1_COPY_7200_9200 (1-2001) x I38022 (1-196)

```
Qy       1587 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1646
              |||||||||      |||:::|||   |||||||||:::|||:::      |||||||||
Db          3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22

Qy       1647 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1706
              ||||||||||:::|||||||||              ||||||   |||||||||||    |||
Db         23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42

Qy       1707 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1766
              ||||||||||||||   ||||||   |||||||||||||||||||||:::||||||
Db         43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62

Qy       1767 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 1826
                 |||   |||||||||   |||:::   ||||||:::||||||:::      ||||||
Db         63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGln------AsnLys 80

Qy       1827 ATAAAGGATCGGAG-------AGAAACAAAACTAATAAGATT------------------ 1861
                 |||:::            ||||||   |||   |||
Db         81 ThrLysAsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100

Qy       1862 ---CCTGAAGGTAAGCAGAGA 1879
                 |||||||||:::::::::
Db        101 GlyProGluGlyGluGluLys 107
```

RESULT 9
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	1.14e-18	Length:	597
Score:	259.50	Matches:	107
Percent Similarity:	40.6%	Conservative:	18
Best Local Similarity:	34.7%	Mismatches:	75
Query Match:	7.4%	Indels:	110
DB:	4	Gaps:	10

US-10-607-806-1_COPY_7200_9200 (1-2001) x E40201 (1-597)

```
Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGT---CTTGCTCTGTCACCCAGGCATGAGCAC 460
              |||||||||                  |||   |||   |||   ||||||   |||:::
Db        302 PhePhePhe------------***AspGlyValLeuLeuLeuLeuProArgProGluTyr 317

Qy        461 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG--- 517
                 |||                                  :::|||   :::   ||||||
Db        318 AspGlyThr--------------------------IleSerProGlnThrLeuLeuPro 328

Qy        518 --------------------ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCAC 556
                                   |||||||||:::|||||||||   |||   :::
Db        329 GlyPheLysGlnPheSerProValSerAlaSerArgValAlaGlyIleThrGlyMetArg 348

Qy        557 CACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCC 616
              |||||||||   ||||||          |||      |||   |||   :::   |||
```

```
Db        349 HisHisAlaGlnLeuIle---LeuTyrPhe******ArgTrpGlyPheSerMetLeuVal 367

Qy        617 AGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCT-CCTACCTTGGCATCCCAAAGTGCT 675
              :::||||||   ||||||      |||  |||  |||   |||  ||||||||||||
Db        368 LysLeuValSerAsnSerGlnProGlnValIleArgProAlaLeuAsnSerGlnSerAla 387

Qy        676 GGGATTACAGGGGTGAGCCACCATGTG------------------------------- 702
              ||||||||||||:::|||:::|||
Db        388 GlyIleThrGlyMetSerTyrHisThrTrp*********PhePhePhePheGluThrGlu 407

Qy        703 ----------------------------------------------CGGCTACTTATTTCT 717
                                                           |||   |||
Db        408 PheCysSerCysCysProGlyArgSerThrMetAlaGlnSerHisArgLysLeuCysPhe 427

Qy        718 TTACATTCC---------------------------------------------------- 726
              |||   |||
Db        428 LeuGlySerSerAsnSerLeuLeuSerGlnProProGlu***LeuGlyLeuGlnAlaCys 447

Qy        727 ------------------------ATCTTT---------------------------- 732
                                       ||||||
Db        448 AlaThrThrProSer***PheTyrIlePheSerArgAspGlyValSerProCysTrpSer 467

Qy        733 ------------------------------------------------CCAATAGAATGT 744
                                                             |||
Db        468 SerTrpSerGlnThrProAsnLeuArg***SerAlaProPro***ThrProLysAlaLeu 487

Qy        745 AAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCT 804
                 :::                |||                |||||||||       |||
Db        488 GlyLeuGlnAla***AlaThrThrProGly*********PhePhePheLeuArgArgSer 507

Qy        805 CACTTCATCACCTCAACCTCC-----------GTTCAGCTCACTGCAACCTCTGCCTCC 852
                  :::   :::   :::                |||||||||      ||||||
Db        508 PheAlaLeuValAlaGlnAlaGlyValArgTrpHisAsnLeuThrAlaAsnPheAlaSer 527

Qy        853 CGGGTTCAAGYG---ATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 909
                 ||||||      ::: ||||||||||||||||   |||||||||:::|||:::   ||
Db        528 TrpValGlnAlaIleLeu-SerCysLeuSerLeuProSerSerTrpAspTyrArgHisAl 547

Qy        910 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 969
              |||||||      ||||||  |||:::||||||   ||||||||||||   ||||||
Db        547 aProProArgProAlaAsn--PheIlePheLeuValGluMetGlyPheLeuHisValGly 566

Qy        970 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGC-CTCAGTCTCCCAAAGTGC 1028
              ||||||||||||||||||   ||||||  |||  |||  |||  |||||||||
Db        567 GlnAlaGlyLeuLysLeuProThrSerGlyAspProProArgLeuGluLeuProLysArg 586

Qy       1029 TGGAATTATAGGCGTGAG 1046
              |||:::||||||   |||
Db        587 TrpAspTyrArgHisGlu 592
```

RESULT 10
PSHU
phospholipase A2 (EC 3.1.1.4) IB precursor [validated] - human
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Homo sapiens (man)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: C25793; A25793; A00733; A21566
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986
A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: C25793
A;Molecule type: DNA
A;Residues: 1-148 <SEI2>
A;Cross-references: UNIPROT:P04054; UNIPARC:UPI0000048FD9; GB:M22970
A;Accession: A25793
A;Molecule type: mRNA
A;Residues: 1-148 <SEI>
A;Cross-references: UNIPARC:UPI0000048FD9; GB:M21055
A;Experimental source: lung
R;Grataroli, R.; Dijkman, R.; Dutilh, C.E.; Van der Ouderaa, F.; De Haas, G.H.; Figarell
Eur. J. Biochem. 122, 111-117, 1982
A;Title: Studies on prophospholipase A2 and its enzyme from human pancreatic juice. Cata
A;Reference number: A91114; MUID:82138816; PMID:7060561
A;Accession: A00733
A;Molecule type: protein
A;Residues: 16-22 <VER1>
A;Cross-references: UNIPARC:UPI000014C3BF
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A21566
A;Molecule type: protein
A;Residues: 23-143,'YS',146-147 <VER2>
A;Cross-references: UNIPARC:UPI0000172758
C;Genetics:
A;Gene: GDB:PLA2G1B; PLA2A; PLA2; P
A;Cross-references: GDB:120715; OMIM:172410
A;Map position: 12q23-12qter
A;Introns: 12/1; 65/2; 108/1
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; me
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <ACT>
F;23-148/Product: phospholipase A2 IB #status experimental <MAT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:

Pred. No.:	8.01e-18	Length:	148
Score:	251.00	Matches:	43
Percent Similarity:	95.7%	Conservative:	2
Best Local Similarity:	91.5%	Mismatches:	2
Query Match:	7.1%	Indels:	0
DB:	1	Gaps:	0

US-10-607-806-1_COPY_7200_9200 (1-2001) x PSHU (1-148)

```
Qy         22 CTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGC 81
              |||   :::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 LeuAspLysCysCysGlnThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSer 82

Qy         82 TGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCG 141
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db         83 CysLysPheLeuLeuAspAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySer 102

Qy        142 GCAATCACCTGTAGCAGTAGG 162
              ||||||||||||||||||:::
Db        103 AlaIleThrCysSerSerLys 109
```

RESULT 11
B40201
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC9
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

.

Alignment Scores:

Pred. No.:	1.57e-17	Length:	301
Score:	248.50	Matches:	80
Percent Similarity:	43.7%	Conservative:	17
Best Local Similarity:	36.0%	Mismatches:	74
Query Match:	7.1%	Indels:	52
DB:	4	Gaps:	8

US-10-607-806-1_COPY_7200_9200 (1-2001) x B40201 (1-301)

```
Qy       1061 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1002
Db          1 GlyArgAlaTrpTrpLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGly 20

Qy       1001 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 942
Db            LeuLeuGluThrArgSerSerArgProAlaTrpAlaThr***GlnAspProSerSerIle 40

Qy        941 AAAAATACAAAAAATTA----------------GCCAAGCATGGTGGTGCACGCTTGTAA 898
Db            LysIleLys***Asn***LeuGly*********AlaGlyHisGlyGly***HisLeu*** 60

Qy        897 TTCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGC 838
Db            SerGlnHisPheGlyMetLeuLys***GluAspCysLeuArgProGly------------ 76

Qy        837 AGTGAGCTGAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAG 778
Db            ------ValGlnAspGlnLeuGlyGlnHisSerLysThrProAlaLeu***Lys***Asn 94

Qy        777 AAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAA 718
Db            LysIleSer---------------------------------TrpVal*********--- 102

Qy        717 AGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAA 658
Db            ------------GlnGlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCys*** 118

Qy        657 GGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCC 598
Db            SerArgArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaArgPro 138

Qy        597 AGTCTCTACAAA---AAAATT--------------------------------TTTTTA 574
Db            GlnLeuTyrLysAsnLysIleLysLeuAlaGly*********HisProAlaAsnPheIle 158

Qy        573 ATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCT------------ACTTGGGAG 526
Db            LeuPheLeu***SerTrpGlyLeuAlaMetLeuProLysLeuValLeuAsnSerTrpSer 178

Qy        525 GCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCA 466
Db            GlnAlaIle-----------------------LeuLeuLeuGlnHisProLysValLeu 190

Qy        465 CCACTG 460
Db        191 GlyLeu 192
```

RESULT 12
B40201
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC9
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	1.41e-16	Length:	301
Score:	239.50	Matches:	81
Percent Similarity:	48.6%	Conservative:	23
Best Local Similarity:	37.9%	Mismatches:	67
Query Match:	6.8%	Indels:	44
DB:	4	Gaps:	10

US-10-607-806-1_COPY_7200_9200 (1-2001) x B40201 (1-301)

```
Qy        467 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 526
Db        113 SerThrLeuGlyCys***SerArgArgIleAla***AspGlnGluPheLysThrSerLeu 132

Qy        527 TCCCAAGTAGCTGGG----------------------------ACTACGAGCGTGCAC 556
Db        133 GlyAsnIleAlaArgProGlnLeuTyrLysAsnLysIleLysLeuAlaGly********* 152

Qy        557 CACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCC 616
Db        153 HisProAlaAsnPheIle------LeuPheLeu***SerTrpGlyLeuAlaMetLeuPro 170

Qy        617 AGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTG 676
Db        171 LysLeuValLeuAsnSerTrpSerGlnAlaIleLeuLeuLeuGlnHisProLysValLeu 190

Qy        677 GGATTACAGGGGTGAGCCACCATG---TGCGGCTACTTATTTCTTTACATTCCATCTTTC 733
Db        191 GlyLeuGlnMetLeuAlaThrMetProCys----------------------------- 200

Qy        734 CAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTG 793
Db        201 ---------*********ThrGln---------LeuIleLeuPheTyr------PheTyr 212

Qy        794 AGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCC 853
Db        213 ArgAlaGlyValLeuLeuCys---CysProSerTrpSer***------------ThrPro 227

Qy        854 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 913
Db        228 GlyLeuLysGlnSerSerTyrPheSerIleProLysCysTrpAspTyrArgCys***Pro 247

Qy        914 CCATGCTTGGCTAA------------------TTTTTTGTATTTTTAGCAGAGATGGGG 954
Db        248 ProCysProAla*********ProSer***PheTyrPheIlePheIle---GluLeuGly 266

Qy        955 TTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCA 1014
Db        267 SerCysTyrValAlaGlnAlaGlyLeuGluLeuLeuValSerSerAsnProProThrSer 286

Qy       1015 GTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCC 1056
Db        287 AlaSerGlnSerAlaGlyIleThrAspValSerHisHisAla 300
```

RESULT 13
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201

```
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              3.12e-16        Length:         579
Score:                  236.50          Matches:        97
Percent Similarity:     41.8%           Conservative:   17
Best Local Similarity:  35.5%           Mismatches:     77
Query Match:            6.7%            Indels:         82
DB:                     4               Gaps:           7

US-10-607-806-1_COPY_7200_9200 (1-2001) x D40201 (1-579)

Qy        397 GTTTGGGTT------------TTTTTTTGT------------------TGTTGT------ 420
              :::|||:::             |||||||||                  ||||||
Db        339 IleTrpLeuIlePheValSerPhePheCysArgGlyGlyValSerThrCysCysProGly 358

Qy        420 ------------------------------------------------------------ 420

Db        359 CysSer***ThrProGluLeuGluHisThrProAlaMetAlaProLysValLeuGlyLeu 378

Qy        421 ------------------------------TTAGAGACAGGGTCTTGCTCTGTCACCCAG 450
                                           :::   ||||||      |||   |||
Db        379 GlnAlaCysSerHisAlaTrp*********PheGlnAlaGlySerTrpTyrValAlaGln 398

Qy        451 GCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGA 510
              |||      |||   |||||||||   ||||||  ||||||:::
Db        399 AlaGlyValGlnArgCysAsnHisGlySerLeu-SerLeuAspPheProGlyGlnValIl 418

Qy        511 TCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGC- 569
               |||||| |||||||||   :::   |||||||||:::|||      |||      |||
Db        418 eLeuLeu-ProGlnProPheGlu***LeuGlyLeuGlnAlaTyrAlaThrArgSerGly* 438

Qy        570 --TAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTT 627
                  |||       ||||||||||||   |||     :::|||   ||||||   |||
Db        438 **PheLeuTyrLeu-PhePheValGluValGlyPheArgHisValAlaGlnAlaValLeu 457

Qy        628 AAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGG 687
              :::|||||:::   ||| |||   ||||||||| :::   ||||||||||||||||||
Db        458 GluLeuLeuSerTrpSer-IleHisLeuProTrp-LeuLeuLysCysTrpAspTyrArgA 477

Qy        688 GTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 747
                  |||   |||   |||
Db        477 rg--AlaAlaMetLeuGly---------------------------------------- 482

Qy        748 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 807
                                                        ||||||   |||
Db        483 ------------------------------------*********PheArgGlnGlyLeuGl 491

Qy        808 TTCATCACCTCAACCTCCGTTCAGCTCACTG------------CAACCTCTGCCTCCCGG 855
                  |||:::      :::                       |||
Db        491 yMetLeuProArgLeuGluTyrArgGlyAlaIleMetAlaHis***AlaLeuThrSerGl 511

Qy        856 GTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACC 915
              ||||||   ||||||   ||||||||||||||||||:::|||:::     ..    ||
Db        511 nValLys***SerSerCysProSerLeuLeuSerSerTrpAspTyrArgLeuMetLeuPr 531

Qy        916 ATGCTTGGCTAATTTT---TTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAG 972
              |   ||||||||||||   ::::::|||      ||||||||||||   ||||||||||
Db        531 oAspLeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPheAspMetLeuProAr 551

Qy        973 GCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGA 1032
                  ||||      ||||||     :::                       |||   |||||||||
Db        551 gLeuPheLeuAsnSer--***AlaGlyAlaTyrThrCysHisGlySer***SerAlaGly 570

Qy        1033 ATTATAGGCGTGAGTCACTGT 1053
               |||   ||||||      |||
Db         571 IleThrGlyValGlnProCys 577

RESULT 14
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked r
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.:              7.09e-16        Length:         100
Score:                  232.50          Matches:        51
Percent Similarity:     75.3%           Conservative:   10
Best Local Similarity:  63.0%           Mismatches:     18
Query Match:            6.6%            Indels:         2
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_7200_9200 (1-2001) x A46010 (1-100)

Qy       1822 TTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGG 1763
              |||   :::|||||||||   ||||||      |||||||:::|||||||||||||||
Db          2 PhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCys 21

Qy       1762 GATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTC 1703
              ||||||||||||||:::     ||| ||||||||||||:::|||   ||||||||||
Db         22 AspLeuGlySerLeuLysSerProPro-ProGlySerSerAspSerProAlaSerAlaSe 41

Qy       1702 CCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGT 1643
              |||||||||||||   |||||||||||||:::|||||||||||||||||||||||||||
Db         41 rArgValAlaGlyIleThrGlyMetHisHisHisThrGlnLeuIlePheValPheLeuVa 61

Qy       1642 AGAGATAGGGTTTCAC---CATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGA 1586
              ||||   |||   |||      :::      :::   |||   :::   |||:::   ::
Db         61 lGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAs 81

Qy       1585 T 1585
              :
Db         81 n 81

RESULT 15
S41044
chromosomal protein - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
C;Accession: S41044
R;Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A;Title: A new chromosomal protein essential for mitotic spindle assembly.
A;Reference number: S41044; MUID:94166884; PMID:8121495
A;Accession: S41044
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-418 <YEO>
A;Cross-references: UNIPROT:P49646; UNIPARC:UPI000013C29A; GB:L26953; NID:g537529; PIDN
C;Superfamily: very very hypothetical protein RMSA-1
C;Keywords: chromosomal protein

Alignment Scores:
```

Pred. No.:	3.08e-15	Length:	418
Score:	227.00	Matches:	84
Percent Similarity:	40.6%	Conservative:	41
Best Local Similarity:	27.3%	Mismatches:	69
Query Match:	6.4%	Indels:	114
DB:	2	Gaps:	12

US-10-607-806-1_COPY_7200_9200 (1-2001) x S41044 (1-418)

```
Qy       1228 TTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTG 1287
Db          4 PheCysLeuIleLeuPheAsnLeu---------PheAsnPheCys---------LeuIle 17

Qy       1288 TTACTGTTTTTTCATTGT-----------------TACTATTTTAGT------------ 1317
Db         18 LeuSerPheLeuThrCysIleAlaTyrProValLeuTyrPhePheSerPheLeuAsnSer 37

Qy       1318 ---------------------------------CACTGAATAATATGG------------ 1332
Db         38 GlnThrArgSerLeuLysLeuPheLeuTyrPheLysIleGluIleTrpSerGlnGlyIle 57

Qy       1333 CTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG--- 1389
Db         58 ValArgPheLeuLeuArgPheGlyLeuThrThrPheLysGluArgPheThrSerLeuIle 77

Qy       1389 ------------------------------------------------------------ 1389
Db         78 LeuLeuAsnAsnMetHisGlnMetIlePheProMetValLysTyrIleSerLysLeuCys 97

Qy       1390 ------------------ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACAC 1431
Db         98 IlePheHisPheTrpHisLeuValLeuMetAspLeuValProArgGlnGlnArgSer-Il 117

Qy       1432 ACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTA 1491
Db        117 eIleThrTyrSerLeuValPheAlaIleIleSerGlnLysLysLysArgGlyIleTyrHi 137

Qy       1492 GAGGATGTTTGTTCATATT---------------------------------------- 1510
Db        137 sLysAsnAsnIleArgIleIleLeuPheLeuProGlnAlaHisGlyArgAspPheTyrVa 157

Qy       1511 ----------------AATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAAC 1554
Db        157 lProIleLeuProPheThrGlnSerTyrValAspTrpGlyArgTrpLeu----------- 173

Qy       1555 CACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGC 1614
Db        174 -------IleTrpGluAlaLysAlaGlyGluSerLeuGluValArgSerSerArgProAl 191

Qy       1615 CTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTG 1674
Db        191 aSerGlnSerArgArgAsnSerValSerThrLysAsnIleLysIleSer----------- 207

Qy       1675 ATGCATGCCTGTA---------------------------------------------- 1687
Db        208 -------ProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGl 225

Qy       1688 ------------GTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGG 1734
Db        225 yValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAspPro-G 245

Qy       1735 GAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGC 1794
Db        245 lyGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThrSer- 264

Qy       1795 GAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAA 1854
Db        265 --GluThrValSerLysLys------LysLysLysLysLysLysLysThrArgValL 282

Qy       1855 TAAGATTCCTGAAGGTAAGC 1874
Db        282 euThrCysIleAsnAlaSer 288
```

Search completed: February 8, 2006, 10:03:49
Job time : 87.6544 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 97.1172 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_7200_9200
Perfect score: 3524
Sequence: 1 ctatctcagctgtcccctccc.....ckctacataaatatctatta 2001

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091819_26843/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_694_@runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	502.5	14.3	591	1	ALU1_HUMAN	P39188 homo sapien
c	2	498	14.1	591	1	ALU1_HUMAN	P39188 homo sapien
	3	479	13.6	593	1	ALU7_HUMAN	P39194 homo sapien
c	4	463	13.1	593	1	ALU7_HUMAN	P39194 homo sapien
	5	443	12.6	593	1	ALU6_HUMAN	P39193 homo sapien
c	6	438.5	12.4	591	1	ALU8_HUMAN	P39195 homo sapien
	7	427.5	12.1	591	1	ALU8_HUMAN	P39195 homo sapien
c	8	424	12.0	593	1	ALU6_HUMAN	P39193 homo sapien
	9	406.5	11.5	587	1	ALU2_HUMAN	P39189 homo sapien
	10	397.5	11.3	375	2	O60448_HUMAN	O60448 homo sapien
	11	396.5	11.3	239	2	Q6ZWA9_HUMAN	Q6zwa9 homo sapien
	12	386	11.0	179	2	Q6ZSI5_HUMAN	Q6zsi5 homo sapien
	13	384.5	10.9	158	2	Q8NAL9_HUMAN	Q8nal9 homo sapien
c	14	382.5	10.9	585	1	ALU5_HUMAN	P39192 homo sapien
	15	381	10.8	241	2	Q6ZTY3_HUMAN	Q6zty3 homo sapien
c	16	377.5	10.7	587	1	ALU2_HUMAN	P39189 homo sapien
	17	372.5	10.6	585	1	ALU5_HUMAN	P39192 homo sapien
	18	372	10.6	200	2	Q6ZUK3_HUMAN	Q6zuk3 homo sapien
	19	369	10.5	587	1	ALU3_HUMAN	P39190 homo sapien
c	20	366.5	10.4	156	2	Q8NBH4_HUMAN	Q8nbh4 homo sapien
	21	365	10.4	603	1	ALU4_HUMAN	P39191 homo sapien
	22	363	10.3	314	2	Q6ZTL7_HUMAN	Q6ztl7 homo sapien
c	23	363	10.3	603	1	ALU4_HUMAN	P39191 homo sapien
c	24	361	10.2	146	2	Q6ZRX6_HUMAN	Q6zrx6 homo sapien
c	25	360.5	10.2	128	2	Q6ZPB2_HUMAN	Q6zpb2 homo sapien
c	26	350.5	9.9	129	2	Q6ZUW8_HUMAN	Q6zuw8 homo sapien
c	27	344.5	9.8	150	2	Q6ZPA0_HUMAN	Q6zpa0 homo sapien
c	28	344	9.8	123	2	Q6ZMQ3_HUMAN	Q6zmq3 homo sapien
c	29	342	9.7	129	2	Q8N9K0_HUMAN	Q8n9k0 homo sapien
	30	339.5	9.6	176	2	Q6ZR92_HUMAN	Q6zr92 homo sapien
c	31	336	9.5	124	2	Q6ZSN0_HUMAN	Q6zsn0 homo sapien
	32	335.5	9.5	170	2	Q8NAI3_HUMAN	Q8nai3 homo sapien
c	33	335.5	9.5	587	1	ALU3_HUMAN	P39190 homo sapien
c	34	334	9.5	164	2	Q6ZUK0_HUMAN	Q6zuk0 homo sapien
c	35	332.5	9.4	137	2	Q6ZS97_HUMAN	Q6zs97 homo sapien
c	36	331.5	9.4	127	2	Q6ZVW2_HUMAN	Q6zvw2 homo sapien
	37	329.5	9.4	157	2	Q8N845_HUMAN	Q8n845 homo sapien
c	38	329	9.3	167	2	Q6ZSN3_HUMAN	Q6zsn3 homo sapien
c	39	328.5	9.3	120	2	Q6ZTA3_HUMAN	Q6zta3 homo sapien
c	40	328	9.3	140	2	Q6ZP99_HUMAN	Q6zp99 homo sapien
c	41	327.5	9.3	115	2	Q9N083_MACFA	Q9n083 macaca fasc
c	42	326.5	9.3	131	2	Q6ZTD3_HUMAN	Q6ztd3 homo sapien
	43	326	9.3	164	2	Q6ZTS0_HUMAN	Q6zts0 homo sapien
	44	326	9.3	334	2	Q6ZR97_HUMAN	Q6zr97 homo sapien
	45	325.5	9.2	171	2	Q9H728_HUMAN	Q9h728 homo sapien

ALIGNMENTS

RESULT 1
ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:                9.46e-42        Length:         591
Score:                    502.50          Matches:        129
Percent Similarity:       48.6%           Conservative:   15
Best Local Similarity:    43.6%           Mismatches:     65
Query Match:              14.3%           Indels:         87
DB:                       1               Gaps:           4

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU1_HUMAN (1-591)

Qy        407 TTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGT 466
              ||||||:::         |||||||||||||||||||||||||||   |||   ||||||
Db        298 PhePheLeu---------ArgGlnGlyLeuAlaLeuSerProArgLeuGluCysSerGly 314

Qy        467 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 526
              |||         ||||||||||||:::|||      |||   |||         ||||||
Db        315 AlaIleThrAlaHisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla 334

Qy        527 TCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTT 586
              |||:::|||||||||||||||         |||||||||   ||||||        ||||
Db        335 SerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArgLeuIlePheVal--PheP 354

Qy        587 TTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCA 646
              ||||||||||||||   :::|||||   ||||||   |||:::||||||   ||||||:
Db        354 heValGluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGluLeuLeuGlySerSerA 374

Qy        647 ATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGG- 705
              ::||||||      |||||||||||||||||||||||||||||||||||      |||
Db        374 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg* 394

Qy        705 ---------------------------------------------------------- 705

Db        394 ********PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAlaValAlaA 414

Qy        706 ----------------------------------------CTACTTATTTCTTTAC 721
                                                     ||||||
Db        414 rgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnProP 434

Qy        722 ATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT- 780
                          ::::::   :::           |||         |||||||||
Db        434 roGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu* 454

Qy        780 ---------------------------------------------------------- 780

Db        454 **ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIleL 474

Qy        781 ----------------------------------------------------------C 781

Db        474 euProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly**** 494

Qy        782 CTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC--------- 832
              ||||||||||||   |||      :::           |||
Db        494 *****PhePheGluThrGlySerArgSerValAlaGlnAlaGlyVal-GlnTrpArgAsp 513

Qy        833 ------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTG 886
                    ||||||||||||   |||||||||   |||       ||||||||||||||||
Db        514 HisGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuPro 533

Qy        887 AGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAG 946
              |||||||||:::|||:::|||||||||||||      |||||| ||||||||||||   ||
Db        534 SerSerTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheCysAr 553

Qy        947 AGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGC 1006
              |||||||||      |||||||||||||||||||:::|||||||   ||||||   |||::
Db        553 gAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSe 573

Qy       1007 CTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAG 1046
              :   |||   |||||||||||||||:::||||||||||||
Db        573 rArgLeuGlyLeuProLysCysTrpAspTyrArgArgGlu 586

RESULT 2
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
```

```
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:              2.73e-41        Length:         591
Score:                  498.00          Matches:        125
Percent Similarity:     61.2%           Conservative:   12
Best Local Similarity:  55.8%           Mismatches:     53
Query Match:            14.1%           Indels:         34
DB:                     1               Gaps:           2

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU1_HUMAN (1-591)

Qy       1061 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1002
              |||:::|||:::   |||||||||:::|||||||||||||||||||   |||||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       1001 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 942
              |||||||||   ||||||   |||||||||||||||||      ||||||:::|||
Db         21 SerLeuGluProArgSerSerArgProAlaTrpAlaThr******AsnProValSerThr 40

Qy        941 AAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGA 882
              ||||||||| |||||||||:::|||||||||      |||:::||||||||:::||
Db         41 LysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValValProAlaThrArgGl 60

Qy        881 GGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCT-------- 830
              ||||   ||||||   |||||||||||||||||||   |||||||||   |||
Db         60 uAlaGluAlaGlyGlySerLeuGluProGlyArgSerArgLeuGln***AlaValIleAl 80

Qy        829 -------GAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 777
                      ||| ||||||      |||   ||||||||||||
Db         80 aProLeuHisSerSerLeu-GlyAspArgAlaArgProCysLeuLysLys--******-- 98

Qy        776 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 717

Db         98 ------------------------------------------------------------ 98

Qy        716 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 657
                      |||      |||||||||   |||   |||||||||||||||   |||||
Db         99 -------***AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 116

Qy        656 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 597
              |   ||||||   |||   |||||||||   |||:::|||||||||   |||:::|||
Db        116 gArgGluAspHisLeuSerProGlyValArgAspGlnProGlyGlnHisSerGluThrPr 136

Qy        596 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 537
              |||||||||  |||          ||||||   ||||||||||||||||      ||||||
Db        136 oSerLeuGln--LysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGln 155

Qy        536 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 477
              |||||||||||||||||      |||   ||||||   ||||||   |||||||||:::
Db        156 LeuLeuGlyArgLeuArgGlnGluAspArgLeuSerProGlyGlyArgGlyCysSerGlu 175

Qy        476 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 417
                      ||||||      ||||||   ||||||:::|||||||||||||||:::
Db        176 Pro***SerArgHisCysThrProAlaTrpAlaThrGluArgAspProValSerLysLys 195

RESULT 3
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
```

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:              2.38e-39      Length:        593
Score:                  479.00        Matches:       117
Percent Similarity:     58.5%         Conservative:  7
Best Local Similarity:  55.2%         Mismatches:    57
Query Match:            13.6%         Indels:        31
DB:                     1             Gaps:          2

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU7_HUMAN (1-593)

Qy        442 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 501
              |||   ||||||      ||||||   :::      ||||||||||||   ||||||
Db        308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProGly 327

Qy        502 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 561
                 |||      :::   |||||||||||||||||||||||||||   |||   ||||||
Db        328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProPro 347

Qy        562 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 621
              |||   ||||||          |||   ||||||   ||         |||||||||
Db        348 ArgProAlaAsn--PheCysIlePheSerArgAspGlyValSerProCysTrpProGlyT 367

Qy        622 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 681
                |||   ||||||      :::   |||:::   ||||||:::||||||||||||||||
Db        367 rpSerArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspT 387

Qy        682 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 741
              |||||   |||||||||      |||         |||
Db        387 yrArgArgGluProProArgProAla*********Phe---------------------- 399

Qy        742 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAG 801
                                                                |||   |||   |
Db        400 ----------------------------------------------Phe***AspGlyV 404

Qy        802 TCTCACTTCATCACCTCAACCTCCGT--------------TCAGCTCACTGCAACCTCT 846
              |||||||          |||                  ||||||||||||||||||
Db        404 alSerLeuLeuLeuProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuA 424

Qy        847 GCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCG 906
                ||||||||||||||||||||||||||   ||||||   ||||||||||||||
Db        424 rgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyA 444

Qy        907 TGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTT 966
                  |||||||||   ||||||  ||||||||||||   |||   |||||||:::||||||
Db        444 laArgHisHisAlaArgLeuIle-PheValPheLeuValGluThrGlyPheHisHisVal 463

Qy        967 GCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGT 1026
                ||||||||||||:::||||||||||||   |||   |||||||||   |||||||||
Db        464 GlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSer 483

Qy       1027 GCTGGAATTATAGGCGTGAGTCACTGTGCC 1056
              |||||||||   ||||||||||||   |||
Db        484 AlaGlyIleThrGlyValSerHisArgAla 493

RESULT 4
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
```

```
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:              1.03e-37      Length:         593
Score:                  463.00        Matches:        118
Percent Similarity:     58.8%         Conservative:   15
Best Local Similarity:  52.2%         Mismatches:     58
Query Match:            13.1%         Indels:         35
DB:                     1             Gaps:           3

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU7_HUMAN (1-593)

Qy       1061 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1002
              |||:::|||:::   |||||||||:::|||||||||||||||||||   |||||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       1001 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 942
              |||   ||||||||||||   ||||||||||||   ||||||   ||||||:::|||
Db         21 SerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThr 40

Qy        941 AAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGA 882
              ||||||||| |||||||||:::|||||||||         ||||||||||||||:::||
Db         41 LysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGl 60

Qy        881 GGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAA----- 827
              ||||   |||||||||||||||||||||||:::|||||||||   ||||||
Db         60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAl 80

Qy        826 ----------CGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 777
                         |||         |||||||||   |||||||||
Db         80 aProLeuHisSerSerLeuGlyAsnLysSerGluThrProSerGlnLys---------- 96

Qy        776 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 717
                                                                      ||
Db         97 ----------------------------------------------------------Ly 97

Qy        716 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 657
              |         |||      |||||||||   |||   |||||||||||||||   |||||
Db         97 s*********AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 117

Qy        656 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 597
              |     |||   |||:::   ||||||   |||:::|||||||||      :::|||
Db        117 gArgValAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrPr 137

Qy        596 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 537
              ||||||     |||         ||||||   |||||||||   |||      ||||||
Db        137 oSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyGlyArgLeu***SerGln 156

Qy        536 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 477
              ||||||||||||||||||      :::   |||:::   ||||||   |||||||||:::
Db        157 LeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGlu 176

Qy        476 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 417
                        ||||||      ||||||   |||                  :::|||
Db        177 ProArgSerArgHisCysThrProAlaTrpAlaThrArg----------AlaLysLeu 192

Qy        416 CAACAAAAAAAA 405
              :::   ||||||
Db        193 ArgLeuLysLys 196

RESULT 5
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
```

```
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              1.14e-35        Length:         593
Score:                  443.00          Matches:        113
Percent Similarity:     56.6%           Conservative:   7
Best Local Similarity:  53.3%           Mismatches:     61
Query Match:            12.6%           Indels:         31
DB:                     1               Gaps:           2

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU6_HUMAN (1-593)

Qy        442 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 501
              |||   ||||||     ||||||   :::     |||   ||||||   ||||||
Db        308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProProGly 327

Qy        502 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 561
                 |||      :::   ||||||||||||||||||||||||||      |||||||||
Db        328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProPro 347

Qy        562 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 621
              |||   ||||||          |||   ||||||   |||         |||   |||
Db        348 ArgProAlaAsn--PheCysIlePheSerArgAspGlyValSerProCysTrpSerGlyT 367

Qy        622 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 681
                |||   ||||||      :::   |||:::   ||||||:::||||||||||||||||
Db        367 rpSerArgThrProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrpAspT 387

Qy        682 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 741
              |||||   |||||||||      |||         |||
Db        387 yrArgArgGluProProArgProAla*********Phe---------------------- 399

Qy        742 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAG 801
                                                                |||   |||   |
Db        400 ----------------------------------------------Phe***AspGlyV 404

Qy        802 TCTCACTTCATCACCTCAACCTCCGT--------------TCAGCTCACTGCAACCTCT 846
              |||||||           |||                 |||||||||   ||||||
Db        404 alSerLeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuA 424

Qy        847 GCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCG 906
              ||||||||||||||||||||||||||||   ||||||   ||||||||||||||   :
Db        424 rgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyM 444

Qy        907 TGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTT 966
              ::   |||||||||   ||||||||||||||   |||   ||||||   ||||||
Db        444 etArgHisHisAlaArgLeuIle-PheValPheLeuValGluThrGlyPheLeuHisVal 463

Qy        967 GCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGT 1026
                ||||||||||||:::|||   ||||||   |||   |||||||||   |||||||||
Db        464 GlyGlnAlaGlyLeuGluLeuProThrSerGlyAspProProAlaSerAlaSerGlnSer 483

Qy       1027 GCTGGAATTATAGGCGTGAGTCACTGTGCC 1056
              |||||||||   ||||||||||||   |||
Db        484 AlaGlyIleThrGlyValSerHisArgAla 493

RESULT 6
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
```

```
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:              3.27e-35       Length:         591
Score:                  438.50         Matches:        120
Percent Similarity:     58.0%          Conservative:   14
Best Local Similarity:  51.9%          Mismatches:     59
Query Match:            12.4%          Indels:         38
DB:                     1              Gaps:           4

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU8_HUMAN (1-591)

Qy       1062 CGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAG 1003
              |||||||||   |||   ||||||   |||   |||||||||          ||| |||||
Db          4 ArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeu---------Gly-GlyAr 20

Qy       1002 ATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGC 943
              ||||   ||||||||||||   ||||||||||||   ||||||   ||||||:::|||
Db         20 gSerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerTh 40

Qy        942 TAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGG 883
              ||||||||| |||||||||:::|||||||||          ||||||||||||||:::|
Db         40 rLysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgG 60

Qy        882 AGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAA---- 827
              |||||   |||||||||||||||||||||||:::|||||||||   ||||||
Db         60 luAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleA 80

Qy        826 -----------CGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAG 778
                         ||| ||||||      ||||||   |||||||||
Db         80 laProLeuHisSerSerLeu-GlyAspArgAlaArgLeuArgLeuLysLys********* 99

Qy        777 AAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAA 718

Db         99 ------------------------------------------------------------ 99

Qy        717 AGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAA 658
                      |||      |||||||||||||||||||||||||||||||
Db        100 AlaGly-GlyGlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrp-Glu---- 117

Qy        657 GGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCC 598
              ||||||   |||:::   ||||||   |||:::|||||||||      :::|||
Db        118 -----GluAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrP 136

Qy        597 AGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCA 538
```

```
                  ||||||     |||         ||||||   ||||||||||||||      |||||
Db        136 roSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGl 155

Qy        537 GCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGA 478
              ||||||||||||||||||      :::   |||:::   ||||||   |||||||||::
Db        155 nLeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGl 175

Qy        477 CCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACA 418
              :           ||||||      ||||||   ||||||:::|||   |||||||||::
Db        175 uProArgSerArgHisCysThrProAlaTrpAlaThrGluArgAspSerValSerLysLy 195

Qy        417 ACAACAAAAAAAAAACCCAA 399
              :          :::   :::
Db        195 s*********ArgAlaGlu 201

RESULT 7
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
```

```
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:              4.35e-34        Length:         591
Score:                  427.50          Matches:        117
Percent Similarity:     56.7%           Conservative:   10
Best Local Similarity:  52.2%           Mismatches:     59
Query Match:            12.1%           Indels:         38
DB:                     1               Gaps:           4

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU8_HUMAN (1-591)

Qy        420 TTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTC 479
              |||   |||   |||   ||||||   ||||||   |||   |||   :::      :::
Db        398 Phe***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArgLeu 417

Qy        480 ACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTG 539
              ||||||   |||   |||       |||   |||       ||||||||||:::   |||
Db        418 ThrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProGlu***Leu 437

Qy        540 GGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGG 599
              ||||||:::|||       ||||||||||             |||       |||   |
Db        438 GlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPhe******--ArgArgG 457

Qy        600 GTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTT 659
              ||   |||   |||||||||||||||   ||||||      |||   |||   |||
Db        457 lyPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeu---- 475

Qy        660 GGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTT 719
                  ||| ||||||||||||||||||   |||||||||      |||:::
Db        476 -----Pro-LysCysTrpAspTyrArgArgGluProProArgProAlaSer-AlaArg** 493

Qy        720 ACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCT 779

Db        493 *******------------------------------------------------------ 495

Qy        780 TCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC------- 832
                  ||||||||||||||||||      :::      |||
Db        496 -------PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArgA 513

Qy        833 --------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 884
                      ||||||||||||   ||||||||||||||   ||||||||||||||||||
Db        513 spLeuGlySerLeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuP 533

Qy        885 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 944
              |||||||||:::|||:::||||||||||||      |||||| ||||||||||||||||
Db        533 roSerSerTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheSer 552

Qy        945 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1004
              |||||||||||   ||||||   ||||||||||||:::||||||||||||:::   |||
Db        553 ArgAspGlyValSerProCysTrpProGlyTrpSerArgThrProAspLeuArg***Ser 572

Qy       1005 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1062
                  ||| ||||||||||||||||||   ||||||||||||   |||   |||
Db        573 ---------Ser-SerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgPro 588

RESULT 8
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
```

```
CC       coding nucleotide sequence.
CC   -----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   -----------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              9.93e-34        Length:         593
Score:                  424.00          Matches:        113
Percent Similarity:     57.1%           Conservative:   16
Best Local Similarity:  50.0%           Mismatches:     62
Query Match:            12.0%           Indels:         35
DB:                     1               Gaps:           3

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU6_HUMAN (1-593)

Qy       1061 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1002
              |||:::|||:::   |||||||||:::||||||||||||||||||   |||||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       1001 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 942
              |||   ||||||   |||   |||||||||       ||||||   ||||||:::|||
Db         21 SerProGluValGlySerSerArgProAla***ProThrTrpArgAsnProValSerThr 40

Qy        941 AAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGA 882
              |||||||| |||||||||:::|||||||||         ||||||||||||||:::||
Db         41 LysAsnThr-LysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGl 60

Qy        881 GGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAA----- 827
              ||||   |||||||||||||||||||||||:::||||||:::   ||||||
Db         60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuArg***AlaGluIleAl 80

Qy        826 ----------CGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 777
                         |||         |||||||||   |||||||||
Db         80 aProLeuHisSerSerLeuGlyAsnLysSerGluThrProSerGlnLys---------- 96

Qy        776 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 717
                                                                      ||
Db         97 ----------------------------------------------------------Ly 97

Qy        716 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 657
              |    |||   |||||||||   |||   ||||||||||||||   |||||
Db         97 s*********AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 117

Qy        656 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 597
              |      |||   |||:::   ||||||   |||:::|||   |||      :::|||
Db        117 gArgAlaAspHisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGluThrPr 137

Qy        596 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 537
              ||||||     |||          ||||||   ||||||||||||          ||||||
Db        137 oSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGln 156

Qy        536 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 477
              ||||||||||||||||||      :::   |||:::   ||||||   ||||||   :::
Db        157 LeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysGlyGlu 176

Qy        476 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 417
                        ||||||      ||||||   |||                 :::|||
Db        177 ProArgSerArgHisCysThrProAlaTrpAlaThrArg-----------AlaLysLeu 192

Qy        416 CAACAAAAAAAA 405
                 ...:::   ||||||
Db        193 ArgLeuLysLys 196

RESULT 9
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   -----------------------------------------------------------------------
```

```
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:              6.08e-32        Length:         587
Score:                  406.50          Matches:        126
Percent Similarity:     47.2%           Conservative:   18
Best Local Similarity:  41.3%           Mismatches:     67
Query Match:            11.5%           Indels:         94
DB:                     1               Gaps:           7

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU2_HUMAN (1-587)

Qy        407 TTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGT 466
              ||||||:::         |||:::   ||||||||||||||||||    |||   ||||||
Db        296 PhePheLeu---------ArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGly 312

Qy        467 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 526
              |||         ||||||   |||   |||      ||||||                ||||||
Db        313 AlaIleSerAlaHisCysLysLeuArgLeuProGlySerArgHisSerProAlaSerAla 332

Qy        527 TCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTT 586
              |||:::|||||||||||||||         |||||||||   ||||||        |||
Db        333 SerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArgLeuIlePheVal--PheL 352

Qy        587 TTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCA 646
                ||||||||||||   :::   |||   |||      |||   ||||||   |||
Db        352 euValGluThrGlyPheHisArgValSerGlnAspGlyLeuAspLeuLeuThrSer***- 371

Qy        647 ATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA---TGTGC 703
               |||:::   ||||||:::|||||||||||||||||||||   |||||||||     ||
Db        372 -SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgProAl 391

Qy        704 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG--------------A 748
              |         ||||||         :::||| :::   |||:::             :
Db        391 a*********PhePhe***AspGlyValSer-Leu---CysArgProGlyTrpSerAlaV 410

Qy        749 TCCACAGAACAGGGATTACTGCC---------------------------------- 771
              ::   |||:::   :::||||||
Db        410 alAlaArgSerArgLeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProG 430

Qy        772 ----------------------------------------------TATTTTCTTCCTT 784
                                                           ||||||   |
Db        430 lnProProGlu***LeuGlyLeuGlnAlaProAlaThrThrProGly***PheLeuTyrP 450

Qy        785 TC-------------------------------------------------------- 786
              ||
Db        450 he******ArgArgGlyPheThrValLeuAlaArgMetValSerIleSer***ProArgA 470

Qy        786 ---------------------------------------------------------- 786

Db        470 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg* 490

Qy        787 --------TTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC------ 832
                      ||||||||||||||||||      :::         |||
Db        490 ********PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArg 509

Qy        833 ---------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 883
                      |||||||||      ||||||||||||      ||||||||||||||||||
Db        510 AspLeuGlySerLeuGlnAlaProProProGlyPheThrProPheSerCysLeuSerLeu 529

Qy        884 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAG 943
              |||||||||:::|||:::|||   ||||||      |||||| ||||||||||||||||
Db        530 ProSerSerTrpAspTyrArgArgProProProArgProAlaAsn-PheCysIlePheSe 549

Qy        944 CAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATC 1003
              ||||||||||||   ||||||   ||||||||||||:::::|||||||||   ||||||
Db        549 rArgAspGlyValSerProCys***ProGlyTrpSerArgSerProAspLeu--ValIle 568

Qy       1004 TGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1060
                    |||:::   ||||||||||||   |||   |||      |||   ||||||
Db        569 ArgProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 587

RESULT 10
O60448_HUMAN
ID   O60448_HUMAN PRELIMINARY;      PRT;   375 AA.
AC   O60448;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Neuronal thread protein AD7c-NTP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RX   MEDLINE=98064067; PubMed=9399956;
RA   Monte S.M., Ghanbari K., Frey W.H., Beheshti I., Averback P.,
RA   Hauser S.L., Ghanbari H.A., Wands J.R.;
RT   "Characterization of the AD7C-NTP cDNA expression in Alzheimer's
RT   disease and measurement of a 41-kD protein in cerebrospinal fluid.";
RL   J. Clin. Invest. 100:3093-3104(1997).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RA   de la Monte S.M., Wands J.R.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF010144; AAC08737.1; -; mRNA.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0006915; P:apoptosis; TAS.
DR   GO; GO:0007417; P:central nervous system development; TAS.
SQ   SEQUENCE   375 AA;  41720 MW;  955443950A5BFEFD CRC64;

Alignment Scores:
Pred. No.:              4.48e-31        Length:         375
Score:                  397.50          Matches:        105
Percent Similarity:     56.0%           Conservative:   12
Best Local Similarity:  50.2%           Mismatches:     60
Query Match:            11.3%           Indels:         32
DB:                     2               Gaps:           4

US-10-607-806-1_COPY_7200_9200 (1-2001) x O60448_HUMAN (1-375)

Qy        397 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 456
              ::::::::::|||              |||                |||||||||||||||
Db        194 IlePheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213

Qy        457 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 516
                ||||||   |||      ||||||||||||   ||||||       |||      :::
Db        214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSer 233

Qy        517 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 576
                  ||||||   ||||||||||||||||   |||   ||||||||||||||||||
Db        234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252

Qy        577 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 636
                  |||   ||||||  :::|||   |||      |||||||||:::|||   |||
Db        253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuIleLeuIleSer-- 271
```

```
Qy        637 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 696
                  :::   |||      ||||||||||||||||||||||||||||||||||||||||
Db        272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 291

Qy        697 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 756
              |||   ||||||:::   :::                        |||
Db        292 HisAlaArgLeuIlePheAsnPhe--------------------Cys------------ 300

Qy        757 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 816
                                                ||||||   |||||||||   :::|||
Db        301 ----------------------------LeuPheGluMetGluSerHisSerValThr 310

Qy        817 TCAA--------------CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 861
              |||              |||       ||||||||||||||||||||||||   ||
Db        311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLy 330

Qy        862 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 921
              |   ||||||||||||||||||   |||||||||:::|||         ||||||
Db        330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisPr 350

Qy        922 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 981
              ||||||  ||||||||||||   |||   ||||||   |||         |||||||||
Db        350 oAlaAsn-PheCysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerG 370

Qy        982 AAACTCCTGACCTCAAG 998
              |||||||||||||:::
Db        370 lnThrProAspLeuArg 375

RESULT 11
Q6ZWA9_HUMAN
ID   Q6ZWA9_HUMAN PRELIMINARY;      PRT;   239 AA.
AC   Q6ZWA9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ41367.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Caudate nucleus;
RA   Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA   Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK123361; BAC85595.1; -; mRNA.
SQ   SEQUENCE   239 AA;  25635 MW;  409CED42F0603D85 CRC64;

Alignment Scores:
Pred. No.:              5.02e-31       Length:         239
Score:                  396.50         Matches:        110
Percent Similarity:     56.0%          Conservative:   25
Best Local Similarity:  45.6%          Mismatches:     68
Query Match:            11.3%          Indels:         38
DB:                     2              Gaps:           4

US-10-607-806-1_COPY_7200_9200 (1-2001) x Q6ZWA9_HUMAN (1-239)

Qy        425 AGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGC 484
              |||      |||   |||||||||||   |||   |||||||||
Db          2 ArgTrpSerLeuThrLeuSerProArgLeuGluCysSerGlyAlaIleSerThrAspSer 21

Qy        485 AGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACT 544
              ||||||:::|||      |||   :::   ||||||||||||||||||||   |||
Db         22 SerLeuHisLeuProGlySerSerAsnSerLeuThrSerAlaSerGlnValGlyGlyIle 41

Qy        545 ACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTT 604
              . |||         |||||||||   ||||||        |||   ||||||||||||   :
Db         42 ThrGlyAlaCysHisHisAlaArgLeuIlePheVal--PheSerValGluThrGlyPheH 61

Qy        605 ACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCAT 664
              :::::|||||||||||   |||:::||||||   |||::::::   ||||||  ||||
Db         61 isHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerAsnAspLeuProThrSerAlaS 81

Qy        665 CCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATT 724
              |||||:::|||||||||||||||||||||||||||      :::
Db         81 erGlnAsnAlaGlyIleThrGlyValSerHisArgAlaGln------------------- 94

Qy        725 CCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTT 784
                    |||       |||                                       |
Db         95 --------ProGluLeuCys-----------------------------------------p 99

Qy        785 TCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA------------------------- 819
              ||      ::::::   |||   |||:::
Db         99 heIleLeuLysSerPheSerPhePheLeuArgGlnGlyLeuCysHisProGlyTrpSerA 119

Qy        820 -----ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGC 874
                         |||||||||      |||   |||:::   |||   ||||||||||
Db        119 laValAlaGlnSerGlnLeuThrGluAlaSerValSerGlnAlaGlnAlaIleLeuLeuP 139

Qy        875 CTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTG 934
              ||:::|||||||||      |||   |||   ::::::||||||||||||||||||||
Db        139 roGlnProProGlu--TyrGlySerThrGlyMetTyrHisHisAlaTrpLeuIlePheTr 158

Qy        935 TATTT---TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGA 991
                  |||         |||   :::::::::||||||||||||||:::||||||
Db        158 p-GlyGlyLeuValAsnLysGlySerHisTyrIleAlaGlnAlaGlyLeuGluLeuLeuA 178

Qy        992 CCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACT 1051
              ||||||:::   ||||||||||||||||   ||||||||||   :::|||||||
Db        178 snSerSerAsnProProAlaSerValSerGlnLysAlaGlyIleIleAspIleSerHisC 198

Qy       1052 GTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGA 1104
              |||||   ||||||          ::::::         |||:::   |||
Db        198 ysAlaGlnProIleLeuGluSerLeuMetPheIleProAsnProLysGlyArg 215

RESULT 12
Q6ZSI5_HUMAN
ID   Q6ZSI5_HUMAN PRELIMINARY;      PRT;   179 AA.
AC   Q6ZSI5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ45499.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Thalamus;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK127408; BAC86963.1; -; mRNA.
SQ   SEQUENCE   179 AA;  19118 MW;  52286CCCA5A4458E CRC64;

Alignment Scores:
Pred. No.:              5.49e-30       Length:         179
Score:                  386.00         Matches:        109
Percent Similarity:     60.2%          Conservative:   15
Best Local Similarity:  52.9%          Mismatches:     50
```

```
Query Match:          11.0%      Indels:      33
DB:                   2          Gaps:        6

US-10-607-806-1_COPY_7200_9200 (1-2001) x Q6ZSI5_HUMAN (1-179)

Qy        433 TCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAA 492
Db          3 SerSerSerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProPro 22

Qy        493 CCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGT 552
Db         23 ProSerGlyPheLysArgPheSerCysLeuSerLeuGlnSerSerTrpAspTyrArgHis 42

Qy        553 GCACCACCACGCCTGGCTAATTAAAAAAATTTTTTT---------GTAGAGACTGGGTCT 603
Db         43 AlaProProCysProAlaGly------------PheCysLeuCysValGluThrGlyPhe 58

Qy        604 TACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCA 663
Db         59 HisHisValGlyGlnAlaSerLeuLysLeuLeuThrSerSerAspProProAlaLeuAla 78

Qy        664 TCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACAT 723
Db         79 SerGlnIleAlaGlyIleThrGlyMetIleHisHisAlaGlnProLeuLeuPheLeu--- 97

Qy        724 TCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCT 783
Db         98 ----------------------------ArgGlnGlyValThr------LeuLeuPro 105

Qy        784 TTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA-GCTCACTGCAAC 842
Db        106 --------------------ArgLeuGluCysSerSerAlaIleLeuAlaHisCysIle 118

Qy        843 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 902
Db        119 LeuCysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThr 138

Qy        903 AGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCA 962
Db        139 GlyVal---------CysProLeuIle-PheValPheValValGluMetGlyPheHisHi 155

Qy        963 TGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCA 1022
Db        155 sValAla-ArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProL 175

Qy       1023 AAGTGCTGGAATTA 1036
Db        175 ysValLeuArgLeu 179

RESULT 13
Q8NAL9_HUMAN
ID   Q8NAL9_HUMAN PRELIMINARY;      PRT;   158 AA.
AC   Q8NAL9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ35131.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Placenta;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK092450; BAC03893.1; -; mRNA.
DR   Ensembl; ENSG00000198057; Homo sapiens.
SQ   SEQUENCE   158 AA;  17414 MW;  E151503015F2FE34 CRC64;

Alignment Scores:
Pred. No.:              7.56e-30      Length:          158
Score:                  384.50        Matches:         96
Percent Similarity:     47.2%         Conservative:    6
Best Local Similarity:  44.4%         Mismatches:      44
Query Match:            10.9%         Indels:          71
DB:                     2             Gaps:            3

US-10-607-806-1_COPY_7200_9200 (1-2001) x Q8NAL9_HUMAN (1-158)

Qy        433 TCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAA 492
Db          3 SerHisSerValThrGlnAlaGlyValGlnTrpCysGlyLeuGlySerLeuGlnProLeu 22

Qy        493 CCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGT 552
Db         23 ProProGlyPheLysArgPhePheCysLeuSerLeuLeuSerSerTrpAspTyrArgArg 42

Qy        553 GCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTT 612
Db         43 ValProProProLeuAlaAsn---------------------------------------- 49

Qy        613 GGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGT 672
Db         50 ------PheCysIle---------------------------------------------- 52

Qy        673 GCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTT 732
Db         52 ------------------------------------------------------------ 52

Qy        733 CCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTT 792
Db         53 ----------------------------------------PhePheSerPhePhePhe 58

Qy        793 GAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC--------------TCACT 837
Db         59 GluLysGluSerLeuSerValThrGlnAlaGlyVal-GlnTrpHisAspLeuGlySerLe 78

Qy        838 GCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAA 897
Db         78 uGlnAlaAlaProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAs 98

Qy        898 TTACAAGCGTGCACCACCATGCTTGGCTAA-TTTTTTGTATTTTTAGCAGAGATGGGGTT 956
Db         98 nTyrArgArgProProProCysProAlaAsnPhePheValPheLeuValGluMetGlyPh 118
```

```
Qy        957 TTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGT 1016
Db        118 eCysHisIleGlyGlnAlaGlyLeuGluLeuLeuThrSerSerAspProProThrSerAl 138

Qy       1017 CTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1062
Db        138 aSerGlnSerAlaGlyIleThrGlyMetSerHisArgThrTrpPro 153
```

RESULT 14
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC   into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC   consensus sequences have been constituted that contain all six
CC   frames conceptual translations of each of these classes of Alu
CC   repeats.
CC -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC   codon, 'XXX' is used to separate the various translation phases.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC   pollution of protein sequence databases with Alu-derived amino
CC   acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC   primate genomes with an average spacing of 4 kb. Some of them are
CC   actively transcribed by pol III. Normal transcripts may contain
CC   Alu-derived sequences in 5' or 3' untranslated regions. However,
CC   cDNA libraries also contain partial and/or rearranged cDNAs
CC   ligated with Alu-derived sequence in any orientation. Although Alu
CC   elements (especially situated on the complementary strand) have a
CC   great potential to create additional/alternative exons,
CC   consideration should be given to the possibility that the presence
CC   of an Alu in an open reading frame may have resulted from a
CC   cloning artifact or may be due to misinterpretation of sequencing
CC   data. This point has been overlooked on several occasions, with
CC   the consequence of erroneous Alu-derived amino acid sequences
CC   being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC   with an Alu-translated entry must be taken as a warning that a
CC   part of Alu repeat may have been artifactually included in the
CC   coding nucleotide sequence.
CC ---
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC ---
DR EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW Hypothetical protein.
FT REGION 1 95 Frame-1.
FT REGION 99 193 Frame-2.
FT REGION 197 291 Frame-3.
FT REGION 295 389 Frame-4.
FT REGION 393 487 Frame-5.
FT REGION 491 585 Frame-6.
SQ SEQUENCE 585 AA; 63994 MW; 46EE8C4F493650A7 CRC64;

Alignment Scores:

Pred. No.:	1.72e-29	Length:	585
Score:	382.50	Matches:	108
Percent Similarity:	56.8%	Conservative:	18
Best Local Similarity:	48.6%	Mismatches:	65
Query Match:	10.9%	Indels:	31
DB:	1	Gaps:	2

US-10-607-806-1_COPY_7200_9200 (1-2001) x ALU5_HUMAN (1-585)

```
Qy       1060 GCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGAT 1001
Db         99 AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAsp 118

Qy       1000 CACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTA 941
Db        119 His--GluValLysArgSerArgProSerTrpProThrTrp***AsnProValSerThrL 138

Qy        940 AAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAG 881
Db        138 ysAsnThr-LysIleSerTrpAlaTrpTrpArgAlaProValValProAlaThrArgGlu 157

Qy        880 GCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGG 821
Db        158 AlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGlu------ 175

Qy        820 TTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCC 761
Db        176 ----------------------------------------------------------Ile 176

Qy        760 CTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAG------TA 707
Db        177 AlaProLeuHisSerSerLeuAlaThrGluArgAspSerValSerLys-Lys******** 196

Qy        706 GCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGAT 647
Db        196 *ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIl 216

Qy        646 TGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAA 587
Db        216 eThrArgSerArgAspArg--AspHisProGlyGlnHisGlyGluThrProSerLeu--L 235

Qy        586 AAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGA 527
Db        235 euLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyA 255

Qy        526 GGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGC 467
Db        255 rgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerA 275

Qy        466 ACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 417
Db        275 rgHisCysThrProAlaTrpArgGlnSerGluThrProSerGlnLysLys 291
```

```
RESULT 15
Q6ZTY3_HUMAN
ID   Q6ZTY3_HUMAN PRELIMINARY;      PRT;   241 AA.
AC   Q6ZTY3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ44120.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Thymus;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK126108; BAC86444.1; -; mRNA.
SQ   SEQUENCE   241 AA;  27428 MW;  823CC90412C3F7AD CRC64;

Alignment Scores:
Pred. No.:              1.93e-29      Length:         241
Score:                  381.00        Matches:        117
Percent Similarity:     53.7%         Conservative:   21
Best Local Similarity:  45.5%         Mismatches:     69
Query Match:            10.8%         Indels:         50
DB:                     2             Gaps:           8

US-10-607-806-1_COPY_7200_9200 (1-2001) x Q6ZTY3_HUMAN (1-241)

Qy        404 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 463
              |||||||||          |||:::  |||  ||||||||||||  |||  |||
Db         14 PhePhePhePhePheLeuArgArgThrLeuThrLeuSerProArgLeuGluCysSer 33

Qy        464 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 523
              |||          ||||||:::|||  |||     |||  |||      ||||||
Db         34 GlyThrIleSerAspHisCysAsnLeuCysLeuProGlySerSerAspSerProThrSer 53

Qy        524 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 583
              ||||||  |||||||||  |||  :::  |||||||||  |||      ||| ||
Db         54 AlaSerHisValAlaGlyIleThrGlyIleArgHisHisAlaArgLeu------Lys-Ph 71

Qy        584 TTTTTGT---------AGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCC 634
              |||||||         |||||| |||:::      ||||||||||||   |||:::||||
Db         71 ePheCysIlePheSerArgAsp-GlyAlaLeuProValGlyGlnAlaGlyLeuGluLeuL 91

Qy        635 TGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCC 694
              ||   |||   :::   |||   ||||||||||||||||||   |||   |||:::||||
Db         91 euThrSerGlyAspLeuProAlaLeuAlaSerGlnSerAlaValIleArgGlyMetSerH 111

Qy        695 AC-----------------CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAA 736
              ||                  ::::::|||                  |||:::|||
Db        111 isArgThrLeuProLysGlyTyrIleArgProTyrLysTyrPheAspSerLeuPhe---- 129

Qy        737 TAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGA 796
                                                 |||      ||||||   ||||
Db        130 -----------------------------------PhePhePhePhePheLeuGluT 137

Qy        797 CAGAG---------------TCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAA 841
              |||||                     :::                 ||||||||||||
Db        137 hrGluPheCysSerCysCysProGlyTrpSerAlaValGlyGlnSerGlnLeuThrAlaT 157

Qy        842 CCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTAC 901
              ||||||||||   ||||||  :::||||||||:::  ||||||  |||  ||||
Db        157 hrSerAlaSerTrpValGlnAlaValLeuLeuProGlnSerProGluTrpLeuGlyLeuG 177

Qy        902 AAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTAGCAGAGATGGGGTTTTACC 961
              |||||         ::::::        ||||||||||||   ||||||||||||:::
Db        177 lnAlaHisAlaArgLeuIle---------PheValPheLeuValGluMetGlyPheHisA 194

Qy        962 ATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCC 1021
                |||   ||||||  ||||||||||||||||||:::          ||||||||||
Db        194 rgValGlyGlnAlaAspLeuLysLeuLeuThrSerAla--His------LeuSerLeuPr 211

Qy       1022 AAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTT 1081
              ||||||||||:::||||||||||||           |||  |||      ::::::::
Db        211 oLysCysTrpAspTyrArgArgGluProProArgProAlaLysTyrPheTyrTyrThrMe 231

Qy       1082 ATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTA 1124
                                                 |||||||||
                                                 :
Db        231 t--------------------------------HisIleVal 234

Search completed: February  8, 2006, 09:43:34
Job time : 498.586 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47 ; Search time 25.495 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 1640
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092016_27646/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_605_@runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	455	27.7	396	4	AAU30455	Aau30455 Novel hum
c	2	427	26.0	324	4	AAU29573	Aau29573 Novel hum
c	3	427	26.0	324	4	AAU29579	Aau29579 Novel hum
c	4	406.5	24.8	397	2	AAR95913	Aar95913 Neural th
c	5	393.5	24.0	433	7	ADC87285	Adc87285 Human GPC
c	6	392.5	23.9	241	4	ABB11464	Abb11464 Human neu
c	7	391.5	23.9	179	8	ADR09211	Adr09211 Human pro
c	8	391	23.8	304	4	AAU30391	Aau30391 Novel hum
c	9	389.5	23.8	375	3	AAB01399	Aab01399 Neuron-as
c	10	389.5	23.8	375	5	ABB81538	Abb81538 Neural th
c	11	389.5	23.8	375	5	AAE29142	Aae29142 AD7c-NTP
c	12	389.5	23.8	375	6	ABR63268	Abr63268 AD7c-NTP
c	13	389.5	23.8	375	6	ADA84017	Ada84017 Human POM
c	14	389.5	23.8	375	6	ABU03024	Abu03024 Human neu
c	15	389.5	23.8	375	6	ABB99774	Abb99774 Amino aci
c	16	389.5	23.8	375	6	ABJ19445	Abj19445 AD7C-neur
c	17	389.5	23.8	375	7	ADB37642	Adb37642 Human neu
c	18	389.5	23.8	375	8	ADR14409	Adr14409 Human NF-
c	19	389.5	23.8	375	9	AEA30020	Aea30020 Human neu
c	20	387.5	23.6	332	5	ADK34410	Adk34410 Novel hum
c	21	387	23.6	411	4	ABG08428	Abg08428 Novel hum
c	22	386	23.5	290	4	ABG21913	Abg21913 Novel hum
c	23	380	23.2	382	4	AAU32610	Aau32610 Novel hum
c	24	380	23.2	382	4	AAU31818	Aau31818 Novel hum
c	25	380	23.2	382	4	AAU32707	Aau32707 Novel hum
c	26	377	23.0	215	4	AAU31513	Aau31513 Novel hum
c	27	376	22.9	302	5	ADK36936	Adk36936 Novel hum
c	28	375	22.9	210	4	AAU31810	Aau31810 Novel hum
c	29	368	22.4	164	8	ADQ67004	Adq67004 Novel hum
c	30	366.5	22.3	156	7	ADB63830	Adb63830 Human pro
c	31	365.5	22.3	381	4	AAU30235	Aau30235 Novel hum
	32	365	22.3	85	4	AAO11283	Aao11283 Human pol
c	33	364.5	22.2	129	5	AAO16534	Aao16534 Phosphoen
c	34	362.5	22.1	117	5	AAU11247	Aau11247 Human neu
c	35	361	22.0	146	8	ADR09646	Adr09646 Human pro
c	36	360.5	22.0	196	7	ADE09985	Ade09985 Novel pro
c	37	359	21.9	239	8	ADQ65417	Adq65417 Novel hum
c	38	357.5	21.8	673	6	ADA84009	Ada84009 Human POM
c	39	355.5	21.7	175	4	AAU31782	Aau31782 Novel hum
c	40	355.5	21.7	296	4	AAU33304	Aau33304 Novel hum
c	41	354.5	21.6	291	5	ADK36986	Adk36986 Novel hum
c	42	354	21.6	449	5	ABJ01057	Abj01057 Ovary cel
c	43	352	21.5	110	5	ABP41552	Abp41552 Human ova
c	44	351	21.4	241	6	ABP96219	Abp96219 Human nuc
c	45	350.5	21.4	129	8	ADQ65903	Adq65903 Novel hum

ALIGNMENTS

RESULT 1
AAU30455
ID AAU30455 standard; protein; 396 AA.
XX
AC AAU30455;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #946.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.

```
XX
PS   Claim 20; Page 295; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 396 AA;

Alignment Scores:
Pred. No.:                3.48e-47      Length:          396
Score:                    455.00        Matches:         137
Percent Similarity:       44.1%         Conservative:    23
Best Local Similarity:    37.7%         Mismatches:      96
Query Match:              27.7%         Indels:          108
DB:                       4             Gaps:            12

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAU30455 (1-396)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||   |||||||||   ||||||||||||
Db         50 AspSerProAlaSerAlaSerGlnValAlaGlyIleThrGlyAlaHisHisHisAlaGln 69

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
              ||||||||||||||||||:::|||         |||||||||   ||||||   ||||||
Db         70 LeuIlePheValPheLeuValGluThrSerSerHisHisValGlyGlnAlaSerLeuLys 89

Qy        781 CTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGAT-ACAGGCGTG 723
              |||   ||||||||||||||   |||   |||||||||:::||||||   ||||||:::
Db         90 LeuLeuThrSerSerAspProLeuAlaLeuAlaSerGlnAsnAlaGlyIleThrGlyMet 109

Qy        722 AGCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGG--- 666
              |||||||||      |||:::        ::::::|||               |||
Db        110 SerHisHisThrTrpProAsn---His***IleGluHisPhe---------***TrpVal 125

Qy        665 AACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGG 606
              |||   :::   |||                    :::     |||
Db        126 AsnCysMetValTrpGlu---------------HisTyrLeuAsnLysAlaIleLeuLys 140

Qy        605 AAATTTTCCCAAATAAAA---AAAGCTAGTGTATGCCCCAAA----------------- 567
                 :::   :::||||||   :::           ||||||
Db        141 PheTyrLysGluIleLysSerArgArgAspProThrProLysAlaTyrLeuIleTrpGly 160

Qy        566 ----------------------AATCTGGAGGAA--------------TATGTAACT 546
                                   |||   |||:::               ::::::|||
Db        161 Arg***Cys***ThrValLeu***AsnLysGluGlnValThrAspThrLysPheLeuThr 180

Qy        545 GGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGT 486
                      ||||||          |||                ::::
Db        181 ***LeuPheLysGlyValGlyArgMet***LysPheThrValProAlaProGluValVal 200

Qy        485 AGAGTT--------------------TGAAGTTGCATTACT----------------- 465
              ||||||                   |||:::
Db        201 ArgValSer***LeuTyrLeuAla***AspLysCysLeuValIleIleValAlaMetLeu 220

Qy        465 ------------------------------------------------------------ 465

Db        221 MetAspArgAsnValIlePheLeuLysGlyProLeuMetGlyPheValGly***LeuCys 240

Qy        465 ------------------------------------------------------------ 465

Db        241 SerGlyPheGlnGlyLysGluTrpPheLeuPheGluSerTyrProSerLeuProThrSer 260

Qy        464 ----------------------------TTTGGCTTGTTTACTTTTATAACCAAAATA 435
                                         :::   |||        :::||||||
Db        261 ProThrThrLeuSerProAlaLeuProLysTyrPheLeu***ValAsnValThrLysAla 280

Qy        434 AACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCT 375
              ||||||    ||||||                                  |||:::   ::
Db        281 AsnAsn-PheLeuGln***SerIle-----------------ProPheTyrIleMetVa 294

Qy        374 GCTTACCTTCAGGAATCTTATTAGT-----------------TTTGTTTCTCTCCGATC 333
              :|||||||||   :::      |||                 |||   |||
Db        294 lLeuThrPheProSerThr***SerGlnTyrThrAsnPhePhePhePheSerPheHisPh 314

Qy        332 CTTTATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGT 273
              |||       |||       |||||||||   ||||||      |||   ||||||||||
Db        314 ePhePhePhePhePheCysPhePheGluThrGluSerHisSerValThrGlnAlaGlyVa 334

Qy        272 GCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTG 213
              |||||||   ||||||||||||||   ||||||||||||||||||:::   ||||||||
Db        334 lGlnTrp***AspLeuGlySerLeu***ProLeuProProArgPheGln***PheSerCy 354

Qy        212 CCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTG 153
              ||||||||||   |||||||||||||||||||||||||||   |||   ||||||||||||
Db        354 sLeuSerLeuLeuSerSerTrpAspTyrArgHisAlaProLeuArgProAlaAsnPheCy 374

Qy        152 TATTTTT---AGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCA 96
              |||||||      |||      |||||||||      ||||||||||||   |||
Db        374 sIlePhe******ArgSerGlyValSerProCysTrpProGlyTrpSer***ThrProAs 394

Qy         95 CCTCAAG 89
              |||||||
Db        394 pLeuLys 396

RESULT 2
AAU29573
ID   AAU29573 standard; protein; 324 AA.
XX
AC   AAU29573;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #64.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 167; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
```

```
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 324 AA;

Alignment Scores:
Pred. No.:              1.05e-43      Length:          324
Score:                  427.00        Matches:         117
Percent Similarity:     50.2%         Conservative:    31
Best Local Similarity:  39.7%         Mismatches:      93
Query Match:            26.0%         Indels:          56
DB:                     4             Gaps:            6

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAU29573 (1-324)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::  ||||||  |||  |||||||||  ||||||  |||||||||||||||
Db         62 AspSerSerAlaSerAlaSerArgValAlaGlyIleThrGlyMetCysHisHisAlaGln 81

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
              ||||||||||||||||||:::|||:::  ||||||||||||  ||||||||||||:::
Db         82 LeuIlePheValPheLeuValGluLysGlyPheHisHisValGlyGlnAlaGlyLeuGlu 101

Qy        781 CTCCGGACCTCAAGTGATCCATT-TGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGT 724
              |||      |||  ||||||  |||||||||||||||||||||||||  |||||||||
Db        102 LeuThrAlaSerGlyAspProThrCysLeuGlyLeuProLysCysTrpAspTyrArgArg 121

Qy        723 GAGCCACCATGCCCAACCTAGTTTCTCTACTTT------AAAGCATGTATTTTAAAAATT 670
                  |||     |||  :::|||     ::::::      |||
Db        122 ***AlaThrAlaProGlyLeuPhePhePhePheLeuArgGlnSerPheThrLeuValAla 141

Qy        669 ATGGAACATCCTTAAATGGAATACTACATAGACATTT----------------AGAAA 628
                  ::::::|||           :::                        |||
Db        142 GlnAlaGly-ValGlnTrpArgAspLeuGlySerLeuGlnProProProProProArgPh 161

Qy        627 AATTAGGTATGCCTGC--------------ACATGGAAATTTTCCCAAATAAAAAAAGC 583
              :::::::::|||               ||||||  :::       :::   ||
Db        161 eLysGlnPheSerCysLeuSerLeuProSerThrTrpSerTrp---ValTyrArgHisAl 180

Qy        582 TAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTATCTTTGGGTAGTAGGATT 523
              |     ||||||  |||      :::
Db        180 aProProCysProAlaAsnPheValPhePhe---------------------------- 190

Qy        522 ATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACTTT 463
                                                                      ||
Db        191 ----------------------------------------------------------Ph 191

Qy        462 TGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAA 403
              |     |||  |||:::  :::          :::  :::::::::  |||||
Db        191 ePhePhePhePheLeuValGluLysGlyPheSerMetLeuLeuArgLeuValLeuAs 211

Qy        402 CTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTT---- 347
              |       ::::::          |||       :::|||
Db        211 nSer***ProHisAspProAspProProAlaSerAlaSerGlnSer-AlaGlyIleThrG 231

Qy        346 ----------------------GTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTT 310
                                   :::|||  ::::::      |||  :::||||
Db        231 lyValSerHisHisThrArgProMetSerPheLysAsnIleTyrPhePhePhePheP 251

Qy        309 TTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCAC 250
              |||||  |||||||||  ||||||||||||||||||||||||  ||||||  |||
Db        251 heGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuSerSerA 271

Qy        249 TGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGG 190
              ||||||  ||||||  ||||||:::||||||||||||||||||  ||||||||||
Db        271 rgGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuSerSerSerTrpA 291

Qy        189 ACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTT 130
              |||||||              |||      ||||||||||||  ||||||  ||||
Db        291 spTyrArgArgValProProMetProGly***PheCysIlePheArgArgAspGlyValS 311

Qy        129 CACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAG 89
              |||||          ||||||||||:::|||      ||||||
Db        311 erProArgTrpSerGlyTrpSerGlnThrProAspLeuLys 324
```

```
RESULT 3
AAU29579
ID   AAU29579 standard; protein; 324 AA.
XX
AC   AAU29579;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #70.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 168; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 324 AA;

Alignment Scores:
```

```
Pred. No.:              1.05e-43        Length:         324
Score:                  427.00          Matches:        117
Percent Similarity:     50.2%           Conservative:   31
Best Local Similarity:  39.7%           Mismatches:     93
Query Match:            26.0%           Indels:         56
DB:                     4               Gaps:           6

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAU29579 (1-324)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::   ||||||   |||   |||||||||   ||||||   ||||||||||||||||
Db         62 AspSerSerAlaSerAlaSerArgValAlaGlyIleThrGlyMetCysHisHisAlaGln 81

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
              ||||||||||||||||||:::|||:::   ||||||||||||   ||||||||||||:::
Db         82 LeuIlePheValPheLeuValGluLysGlyPheHisHisValGlyGlnAlaGlyLeuGlu 101

Qy        781 CTCCGGACCTCAAGTGATCCATT-TGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGT 724
              |||       |||   ||||||   |||||||||||||||||||||||   |||||||||
Db        102 LeuThrAlaSerGlyAspProThrCysLeuGlyLeuProLysCysTrpAspTyrArgArg 121

Qy        723 GAGCCACCATGCCCAACCTAGTTTCTCTACTTT------AAAGCATGTATTTTAAAAATT 670
                          |||       |||   :::|||       ::::::      |||
Db        122 ***AlaThrAlaProGlyLeuPhePhePheLeuArgGlnSerPheThrLeuValAla 141

Qy        669 ATGGAACATCCTTAAATGGAATACTACATAGACATTT------------------AGAAA 628
                      ::::::|||             :::                      |||
Db        142 GlnAlaGly-ValGlnTrpArgAspLeuGlySerLeuGlnProProProProProArgPh 161

Qy        627 AATTAGGTATGCCTGC---------------ACATGGAAATTTTCCCAAATAAAAAAAGC 583
                  :::::::::|||               ||||||   :::         :::   ||
Db        161 eLysGlnPheSerCysLeuSerLeuProSerThrTrpSerTrp---ValTyrArgHisAl 180

Qy        582 TAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTATCTTTGGGTAGTAGGATT 523
              |      ||||||   |||       :::
Db        180 aProProCysProAlaAsnPheValPhePhe---------------------------- 190

Qy        522 ATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACTTT 463
                                                                      ||
Db        191 ----------------------------------------------------------Ph 191

Qy        462 TGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAA 403
              |      |||   |||:::   :::           :::   :::::::::   |||||
Db        191 ePhePhePhePheLeuValGluLysGlyPheSerMetLeuLeuArgLeuValLeuAs 211

Qy        402 CTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTT---- 347
              |         ::::::           |||         :::|||
Db        211 nSer***ProHisAspProAspProProAlaSerAlaSerGlnSer-AlaGlyIleThrG 231

Qy        346 ----------------------GTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTT 310
                                     :::|||   ::::::         |||   :::||||
Db        231 lyValSerHisHisThrArgProMetSerPheLysAsnIleTyrPhePhePhePhePheP 251

Qy        309 TTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCAC 250
              |||||   |||||||||   ||||||||||||||||||||||   ||||||   |||
Db        251 heGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuSerSerA 271

Qy        249 TGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGG 190
                ||||||   ||||||   ||||||:::||||||||||||||||||   ||||||||||
Db        271 rgGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuSerSerSerTrpA 291

Qy        189 ACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTT 130
              ||||||||         |||       ||||||||||   ||||||   ||||
Db        291 spTyrArgArgValProProMetProGly***PheCysIlePheArgArgAspGlyValS 311

Qy        129 CACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAG 89
              |||||         |||||||||:::|||      ||||||
Db        311 erProArgTrpSerGlyTrpSerGlnThrProAspLeuLys 324

RESULT 4
AAR95913
ID   AAR95913 standard; protein; 397 AA.
XX
AC   AAR95913;
XX
DT   13-NOV-1996  (first entry)
XX
DE   Neural thread protein.
XX
KW   Neural thread protein; NTP; diagnosis; detection; Alzheimer's disease;
KW   neuroectodermal tumour; malignant astrocytoma; monoclonal antibody;
KW   binding fragment.
XX
OS   Homo sapiens.
XX
PN   WO9615272-A1.
XX
PD   23-MAY-1996.
XX
PF   14-NOV-1995;   95WO-US017111.
XX
PR   14-NOV-1994;   94US-00340426.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   De La Monte S,  Wands JR;
XX
DR   WPI; 1996-259865/26.
DR   N-PSDB; AAT27738.
XX
PT   Detection of neural thread protein in diagnosis of Alzheimer's disease -
PT   also NTP DNA and protein sequences used in gene and anti:sense therapy.
XX
PS   Claim 22; Page 171-172; 238pp; English.
XX
CC   A method for detecting the presence of neural thread protein (NTP) having
CC   a molecular weight of 8, 14, 17, 21, 26 or 42 kD in a human subject
CC   comprises (a) contacting a sample from a human subject that is suspected
CC   of containing the NTP with at least one molecule capable of binding to
CC   the protein; and (b) detecting any of the molecule bound to the protein.
CC   The binding molecule is selected from an antibody free of natural
CC   impurities, a monoclonal antibody or a binding fragment of either of
CC   these. The method may be used for diagnosing the presence of Alzheimer's
CC   disease, neuroectodermal tumours and a malignant astrocytoma in a human
XX
SQ   Sequence 397 AA;

Alignment Scores:
Pred. No.:              4.21e-41        Length:         397
Score:                  406.50          Matches:        144
Percent Similarity:     44.8%           Conservative:   35
Best Local Similarity:  36.0%           Mismatches:     78
Query Match:            24.8%           Indels:         144
DB:                     2               Gaps:           13

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAR95913 (1-397)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGC-TGGGACTACAGGTGCTTGCCACCATGCCCA 843
              |||:::|||||||||   |||   |||   ||||||||||||   ::: |||||||||::
Db         29 AspSerProAlaSerAlaSerProValGlyTrpAspTyrArgHisVal-HisHisAlaAr 48

Qy        842 GCTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCT 786
              :||||||   :::   ||||||:::|||       |||   ||||||   ||||||||||
Db         48 gLeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLe 68

Qy        785 CAAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGC 726
              |:::|||||| ||||||:::|||         ||||||||||||||||| |||||||||
Db         68 uGluLeuArg-ProGlnMetIleLeuProSerArgProProLysValLeu-AspThrGly 87

Qy        725 GTGAGCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGG 666
              :::||| |||||||||                                          ||
Db         88 LeuSer-ThrMetPro------------------------------------------Gl 93

Qy        665 AACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGG 606
              |      |||                 |||:::         |||      |||:::
Db         93 yLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuMetCysProSerTrpSe 113
```

```
Qy        605 AAATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACT 546
Db        113 rProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysTrp-------------- 128

Qy        545 GGTTATCTTTGGGTAGTAGGATTATAGA------TGCCTTCCA---CTTTCCATAATAGA 495
Db        129 ------------------AspTyrArgArgAlaAlaValProGlyLeuPheIleLeu-P 142

Qy        494 TATTTATG---------------------------------------------------- 487
Db        142 hePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpCysAspHisS 162

Qy        486 -------------------------TAGAGTTTGAAGTTG------------------- 472
Db        162 erSerLeuGlnProSerThrLeuArgSerSerIleLeuLeuProGlnProProLysValA 182

Qy        471 --------------------CATTACTTTTGGCTT------GTTTACTTTTATAACCAAA 438
Db        182 laGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIlePheIlePheAsnPheL 202

Qy        437 ATAAACAATAAGATTACAGTTAAAATGCA------------------------------ 409
Db        202 euArgGln---------SerLeuAsnSerValThrGlnAlaGlyValGlnTrpArgAsnL 219

Qy        408 ----------------------------TTTAAACTTTATTACATA------------- 391
Db        219 euGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerCysProSerLeuLeuS 239

Qy        390 -------------------------------------TAATTTACGTATCTC------- 376
Db        239 erSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheValPheLeuValGluM 259

Qy        375 ----------------------------------------------------TGCTTAC 369
Db        259 etGlyPheHisHisValArgGlnValAspAlaArgSerLeuAspLeuValIleCysLeuP 279

Qy        368 CTTCA---------------GGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTT 324
Db        279 roArgProProLysValLeuGlyLeuGlnAspValThrProThrAlaArgProIlePheA 299

Qy        323 ATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTG 264
Db        299 snPheCysLeu--PheGluMetGluSerHisSerValThrGlnAlaGlyValGlnTrpPr 318

Qy        263 GGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCT 204
Db        318 oAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSerCysLeuSerLe 338

Qy        203 CCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAG 144
Db        338 uProSerSerTrpAspTyrGlyHisLeu-HisHisThrProLeuIlePheValPheSerL 358

Qy        143 TAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATC 84
Db        358 euGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSerGlyAspP 378

Qy         83 CACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGT 40
Db        378 roProAlaSerAlaPheGlnSerAlaGlyIleThrGlyValThr 392
```

RESULT 5
ADC87285
ID ADC87285 standard; protein; 433 AA.
XX
AC ADC87285;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1738.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR N-PSDB; ADC87284.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 2; SEQ ID NO 1738; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ Sequence 433 AA;

Alignment Scores:

Pred. No.:	1.86e-39	Length:	433
Score:	393.50	Matches:	128
Percent Similarity:	43.8%	Conservative:	32
Best Local Similarity:	35.1%	Mismatches:	83
Query Match:	24.0%	Indels:	122
DB:	7	Gaps:	10

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ADC87285 (1-433)

```
Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
Db        100 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisValProProArgProAla 119

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
Db        120 AsnPheCysIlePheSerArgAspGly-PheHisHisValGlyGlnLeuGlyLeuLysLe 139

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGT---G 723
Db        139 uLeuThrSerSerAspGlnProAlaLeuAlaSerGlnSerAlaGlyIle-AsnArgCysG 159

Qy        722 AGCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAAC 663
Db        159 luProProArgProAlaGluValIleSerPhe-SerHisLeuPhe--------------- 173

Qy        662 ATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAA 603
Db        173 ------------------------------------------------------------ 173

Qy        602 TTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGT 543
Db        174 ----------------------------------------------------SerLeuGly 176

Qy        542 TATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTA------ 489
Db        177 PheSerLeuIlePheProValLeuLeuProGluLysPheProGlnProTyrLeuProIle 196
```

```
Qy        489 ------------------------------------------------------------ 489

Db        197 PheAlaGluIlePheIlePheAlaIleMetLeuLeuIleSerMetValSerLeuGluSer 216

Qy        488 TGTAGAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATA--------------- 444
                  ||||||   |||         |||   :::|||:::::::::
Db        217 ProArgVal---SerLeuAlaMetPheSerIlePheSerTyrLeuTyrAlaAspIleHis 235

Qy        443 ---ACCAAAATAAACAATAAGATT---------ACAGTTAAAATGCATTTAAACTTTATT 396
                         |||   |||         ||||||            :::|||
Db        236 TyrProSerPheGluAsnGlyIleProProGlyThrValAlaLysGlyCysSerPhePhe 255

Qy        395 ACATATAATTTACGTATCTCTGCTTACCTTC---------------------------- 365
                  :::::: ::::::      ||||||   |||
Db        256 SerPhe-LeuPheAlaCysLeuLeuValPheValLeuValPheSerThrAlaGlyLeuSe 275

Qy        364 ------------------------------------------------AGGAATCTT-- 356
                                                             ||||||
Db        275 rLeuSerValTrpHisSerGlyLeuSerAlaSerIleSerAlaAlaAlaTrpAsnLeuCy 295

Qy        355 ------------------------ATTAGTTTTGTTTCTCTCCGATCCTTTATTTTA-- 323
                                     :::|||         |||   |||::::::|||
Db        295 sValProValAlaGluTrpGlyCysValSerAsnCysGlnLeuProSerTrpMetLeuTr 315

Qy        322 ----------------TTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGC 279
                              |||   :::||||||         |||      ||||||||||
Db        315 pTrpSerSerThrPhePhePhePhePheLeuArgTrpSerLeuThrValAlaGlnAl 335

Qy        278 TGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATT 219
                  ||||||||||||   :::|||||||||||||||      ||||||   ||||||   ||
Db        335 aGlyValGlnTrpCysAsnLeuGlySerLeuGlnLeuProProProMetPheLysProPh 355

Qy        218 CTCTTGCC---TCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGC 162
                  |||| :::   ||||||||||||||||||          |||      |||:::      :
Db        355 eSer-SerLeuSerAlaSerArgValAlaMetIleLeuGlyAlaCysHisHisAlaTrpI 375

Qy        161 TAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAAC 102
                  ::|||||||||||||||         ||||||      :::   ||||||||||||:::|
Db        375 leIlePheValPheLeuAlaThrThrGlyPheArgArgValGlyGlnAlaGlyLeuLysL 395

Qy        101 TTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGA 42
                  ||||||||:::   :::||||||   |||      ||||||      :::||||||:::|
Db        395 euLeuThrThrGlyAsnProProAlaLeuValSerGlnSerAlaGlyIleThrGlyValS 415

Qy         41 GTCAC 37
                  |||||
Db        415 erHis 416
```

RESULT 6
ABB11464
ID ABB11464 standard; peptide; 241 AA.
XX
AC ABB11464;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human neuronal thread protein homologue, SEQ ID NO:1834.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA08708.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
PS Claim 20; Page 194; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention
XX
SQ Sequence 241 AA;

Alignment Scores:

Pred. No.:	1.99e-39	Length:	241
Score:	392.50	Matches:	117
Percent Similarity:	45.3%	Conservative:	14
Best Local Similarity:	40.5%	Mismatches:	77
Query Match:	23.9%	Indels:	82
DB:	4	Gaps:	3

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABB11464 (1-241)

```
Qy        898 ACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCTA 839
Db         34 SerProAlaSerAlaSerArgValAlaGlyThrThrGlyThrArgHis***AlaArgLeu 53

Qy        838 ATTTTTGT-ATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
Db         54 IlePheTyrIlePheSerArgAspGlyValSerProCys***ProGlyTrpSer***Ser 73

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGAGC 720
Db         74 ProAspLeu--ValIleArgProProArgLeuProLysCysTrpAspTyrArgArgGluP 93

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATC 660
Db         93 roProArgProAla***PhePheValPhe------------------------------ 102

Qy        659 CTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTT 600
Db        102 ------------------------------------------------------------ 102

Qy        599 TCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTAT 540
Db        103 -------------------LeuValGluGlnGlyPheThrMetLeuAlaArgMetValS 116

Qy        539 CTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTT 480
Db        116 erIleSer***ProGlnCysAsp--LeuProAlaSerVal-------------------- 128

Qy        479 TGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACA 420
Db        129 ------------------------------------------------------SerGl 130

Qy        419 GTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAA 360
Db        130 nAsnAlaGlyIleThrGlyValSerHisCysAlaTrpProCysLeu-------------- 145

Qy        359 TCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGA 300
Db        146 ------------------------HisPheCysPhePheGlyPhePhePheGluMetGl 157

Qy        299 GTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCT 240
Db        157 uSerCysSerValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaPr 177

Qy        239 GCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCA 180
Db        177 oProProGlyPheThrProPheSerCysLeuSerLeuProGlySerTrpAspTyrArgAr 197

Qy        179 TGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTG 120
Db        197 gProProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerProCys** 217

Qy        119 ACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTA 60
Db        217 *ProGlyTrpSerArgSerProAspLeu--ValIleArgProProArgProProLysVal 236

Qy         59 CTGTGGTTACAGGCT 45
Db        237 LeuGlyLeuGlnAla 241
```

```
RESULT 7
ADR09211
ID   ADR09211 standard; protein; 179 AA.
XX
AC   ADR09211;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Human protein useful for treating neurological disease Seq 2717.
XX
KW   human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   N-PSDB; ADR07255.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 2717; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polypeptide is a protein
CC   encoded by a full length human cDNA sequence of the invention. NOTE: This
CC   sequence is not given in the sequence listing of the specification but
CC   can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC   office.
XX
SQ   Sequence 179 AA;
```

```
Alignment Scores:
Pred. No.:              2.37e-39        Length:         179
Score:                  391.50          Matches:        104
Percent Similarity:     41.4%           Conservative:   14
Best Local Similarity:  36.5%           Mismatches:     33
Query Match:            23.9%           Indels:         134
DB:                     8               Gaps:           5

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ADR09211 (1-179)

Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
Db         29 PheSerCysLeuSerLeuGlnSerSerTrpAspTyrArgHisAlaProProCysProAla 48

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
Db         49 GlyPheCysLeuCysValGluThrGly-PheHisHisValGlyGlnAlaSerLeuLysLe 68

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGAGC 720
Db         68 uLeuThrSerSerAspProProAlaLeuAlaSerGlnIleAlaGlyIle---------- 84

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATC 660
```

```
Db         84 ------------------------------------------------------------ 84

Qy        659 CTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTT 600
                                                 |||||| :::|||
Db         85 --------------------------------ThrGlyMet-IleHis---------- 89

Qy        599 TCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTAT 540

Db         89 ------------------------------------------------------------ 89

Qy        539 CTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTT 480

Db         89 ------------------------------------------------------------ 89

Qy        479 TGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACA 420
                  |||
Db         90 --------His---------------------------------------------- 90

Qy        419 GTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAA 360

Db         90 ------------------------------------------------------------ 90

Qy        359 TCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGA 300
                              :::   ||||||            :::||||||||||
Db         91 --------------------AlaGlnProLeu-----------LeuPheLeuArgGlnG 100

Qy        299 GTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCT 240
                 :::   :::   ||||||||||||||||||      |||   |||||||||   ||||
Db        100 lyValThrLeuLeuProArgLeuGluCysSerSerAlaIleLeuAlaHisCysIleLeuC 120

Qy        239 GCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCA 180
              ||||||||||||||||:::|||   ||||||||||||:::||||||||||||||||||:
Db        120 ysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyV 140

Qy        179 TGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTG 120
                 ::              ||||||||||||||:::||||||:::||||||||||||:::
Db        140 al---------CysProLeuIlePheValPheValValGluMetGlyPheHisHisVal- 156

Qy        119 ACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTA 60
                 |||||||||   |||   ||||||||||||||||||||||||      ||||||
Db        157 AlaArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProLysVal 176

Qy         59 CTGTGGTTA 51
              |||   |||
Db        177 LeuArgLeu 179

RESULT 8
AAU30391
ID   AAU30391 standard; protein; 304 AA.
XX
AC   AAU30391;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #882.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 287; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 304 AA;

Alignment Scores:
Pred. No.:              3.34e-39        Length:         304
Score:                  391.00          Matches:        129
Percent Similarity:     50.8%           Conservative:   26
Best Local Similarity:  42.3%           Mismatches:     93
Query Match:            23.8%           Indels:         60
DB:                     4               Gaps:           9

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAU30391 (1-304)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||   ||||||   ||||||||||||
Db         28 AspSerProAlaSerAlaSerGlnValAlaGlyIleThrGlyValCysHisHisAlaPro 47

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
                 |||||||||   |||:::|||      ||||||||||||   ||||||||||||:::
Db         48 ProIlePheValLeuLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGlu 67

Qy        781 CTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGAT-ACAGGCGTG 723
              |||   ||||||||||||      |||   |||||||||||||||||   ||||||
Db         68 LeuLeuThrSerSerAspLeuProAlaTrpAlaSerGlnSerAlaGlyIleThrGlyGly 87

Qy        722 AGCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAAC 663
              ||||||            :::           |||      |||   |||::::::
Db         88 SerHisLeuCysLeuAlaAsnValLysTyrPheLysSerHisPhe---LysIlePheVal 106

Qy        662 ATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGG--- 606
              |||   :::                                          ||||||
Db        107 IleAspGlu-----------------------------------------ThrTrpPhe 112

Qy        605 AAATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACT 546
              :::   :::   :::          |||   |||
Db        113 GlnHisThr***LeuLeuSerLeuSerSerCysAsp***-------------------- 125

Qy        545 GGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGT 486
                          |||                |||                    ||||||
Db        126 ------------Val***ThrSerIle***ProGlnValLeuAsp***AsnValLeuCys 141

Qy        485 AGAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAG 426
                             ::::::      |||      ::::::   ||||||
Db        142 ---------ProAspValSerGln***LeuLeuProTrpLeuIleLysIle--------- 155

Qy        425 ATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTA------------CGTATC 378
```

```
                                    |||:::   |||   |||           :::
Db        156 ----------------------PheLeuLysTyrThrLeuAspCysTrp***LysArg 167

Qy        377 TCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTAT--------- 327
                ||:::|||         |||   :::   |||         |||:::
Db        168 GlnAlaTrpLeu---------TyrTrpTyr***PheGlyThrGluLeuPheGlyCysPro 184

Qy        326 --------------TTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCA 282
                           ::: :::|||                          ||||||||
Db        185 ***GlyArgGlnThrLeuCysPhe-PhePheLeuArgArgSerLeuAspSerValAlaGl 204

Qy        281 GGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCA 222
              ||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db        204 nAlaGlyValGln***CysAspLeuGlySerLeuGlnProLeuProProArgPheLys** 224

Qy        221 ATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGC 162
              ||||||||||||||   |||:::|||||||||||||||   |||         |||||
Db        224 *PheSerCysLeuSerPheProAsnSerTrpAspTyrArgCysAlaProProLeuProAl 244

Qy        161 TAATTTTTGTATTTTTAGTAGAGATAG-GGTTTCACCATATTGACCAGGCTGGTCTC-GA 104
              |||||||||||||||||||||||||   ||||||   :::|||   |||||||||   ||
Db        244 aAsnPheCysIlePheSerArgAspMetGlyPheAlaMetLeuValArgLeuValSerGl 264

Qy        103 ACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTT 44
              ||||||||||||   |||||||||          ||||||      :::|||||::
Db        264 uLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyMe 284

Qy         43 GAGTCACTGTGCC 31
              :|||:::   |||
Db        284 tSerTyrArgAla 288
```

RESULT 9
AAB01399
ID AAB01399 standard; protein; 375 AA.
XX
AC AAB01399;
XX
DT 20-OCT-2000 (first entry)
XX
DE Neuron-associated protein.
XX
KW Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
KW Parkinson's disease; demyelinating disease; meningitis; prion disease;
KW kuru; Creutzfeldt-Jakob disease; neurofibromatosis; cerebral palsy;
KW muscular dystrophy; central nervous system; CNS;
KW peripheral nervous system; PNS; myopathy; schizophrenia;
KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
KW ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW Werner syndrome, trauma; human.
XX
OS Homo sapiens.
XX
PN WO200034477-A2.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US030408.
XX
PR 11-DEC-1998; 98US-00210083.
PR 09-FEB-1999; 99US-0119365P.
PR 16-MAR-1999; 99US-0124687P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-Young J;
PI Yang J, Lu DAM, Azimzai Y;
XX
DR WPI; 2000-423423/36.
XX
PT New human neuron-associated proteins and polynucleotides encoding them,
PT useful for diagnosis, treatment and prevention of cell proliferative
PT disorders including cancer, neuronal and neurological disorders.
XX
PS Disclosure; Page 143-144; 145pp; English.
XX
CC Human neuron-associated proteins (NEUAP) can be used for for treating or
CC preventing a disorder associated with decreased expression or activity of
CC NEUAP. Antagonists of NEUAP are useful for treating or preventing
CC disorder associated with increased expression or activity of NEUAP. NEUAP
CC or their fragments or derivatives are useful for treating neurological
CC disorder such as epilepsy, ischemic cerebrovascular disease, stroke,
CC cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's
CC disease, dementia and Parkinson's disease. NEUAPs are also useful for
CC treating other demyelinating diseases, bacterial and viral meningitis,
CC prion diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC metabolic diseases of the nervous system, neurofibromatosis, other
CC developmental disorders of the central nervous system, cerebral palsy,
CC neuroskeletal disorders, autonomic nervous system disorders, cranial
CC nerve disorders, spinal cord diseases, muscular dystrophy and other
CC neuromuscular disorders, peripheral nervous system disorders, inherited,
CC metabolic, endocrine, and toxic myopathies, mental disorders including
CC mood, anxiety and schizophrenic disorders, a cell proliferative disorder
CC such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
CC cirrhosis, hepatitis, mixed connective tissue disease (MCTD),
CC myelofibrosis, paroxysmal nocturnal haemoglobinuria, cancers of the
CC adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, and an
CC autoimmune/inflammatory disorder such as acquired immunodeficiency
CC syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
CC allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, Werner
CC syndrome, complications of cancer, hemodialysis, and extracorporeal
CC circulation, viral, bacterial, fungal parasitic, protozoal, and
CC helminthic infections, and trauma. This protein was designated g3002527
XX
SQ Sequence 375 AA;

Alignment Scores:

Pred. No.:	5.58e-39	Length:	375
Score:	389.50	Matches:	128
Percent Similarity:	46.2%	Conservative:	40
Best Local Similarity:	35.2%	Mismatches:	88
Query Match:	23.8%	Indels:	108
DB:	3	Gaps:	15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAB01399 (1-375)

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||  ||||||   |||   ||||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
              ||||||   :::   ||||||:::|||      |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
              :::|||   ||||||   ||||||      ||||||||||||||||    ||||||
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
                   |||||||||:::
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                   ||||||:::|||     ||||||                        |||  |
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------ 627
              ||:::                          :::|||:::
Db        128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 --------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
```

```
                          :::|||:::   |||   |||           :::::::::
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                  :::|||:::   |||          |||
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
                |||   |||                   |||   |||::::::   :::
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                  |||         :::|||   :::     :::   :::|||
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ---------------------------------------------TTTGGCTTGTTTACT- 450
                                                          |||   ||||||:::
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ------------------------------------------------------TTTA 446
                                                                     |||:
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
              ::      :::           |||:::         |||      |||:::
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
              :::|||      ||||||          |||:::   ||||||      :::
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
                :::::|||   :::      ||||||:::||||||      |||   |||||||||
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
              |||||||||   :::|||||||||||||||||||||||||||      |||:::||||||
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
              ||||||||||||||||||||||||||||||   |||          ||||||||||||||
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
              |||||||||   |||      ||||||||||||      |||||||||:::|||
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
              |||:::
Db        374 LeuArg 375
```

RESULT 10
ABB81538
ID ABB81538 standard; protein; 375 AA.
XX
AC ABB81538;
XX
DT 02-SEP-2002 (first entry)
XX
DE Neural thread protein (NTP).
XX
KW Neural thread protein; NTP; Harlil peptide; Alzheimer's disease;
KW Down's syndrome; neuroectodermal tumour; astrocytoma; glioblastoma;
KW hypoxia; ischaemia; cerebral infarction.
XX
OS Homo sapiens.
XX
PN WO200234915-A2.
XX
PD 02-MAY-2002.
XX
PF 25-OCT-2001; 2001WO-US042813.
XX
PR 27-OCT-2000; 2000US-00697590.
XX
PA (NYMO-) NYMOX PHARM CORP.
XX
PI Fitzpatrick J, Averback P, Focht MSS, Bibiano R;
XX
DR WPI; 2002-507998/54.
DR N-PSDB; ABN89470.
XX
PT New Harlil peptide sequences of the Neural Thread Protein, useful in
PT therapeutic assays, e.g. as targets for developing drugs for treating
PT Alzheimer's disease, ischemia or cerebral infarction, or for diagnosing
PT these diseases.
XX
PS Example 1; Fig 1; 53pp; English.
XX
CC The present invention describes a neural thread protein (NTP) peptide
CC having an amino acid sequence selected from ABB81511 to ABB81529 and
CC their homologues, which are referred collectively as Harlil peptides (I).
CC (I) have neuroprotective, nootropic, vasotropic and cerebroprotective
CC activities, and can be used in peptide therapy. The Harlil peptide
CC sequences can be used as analogues for NTP in therapeutic or diagnostic
CC assays by replacing NTP with the peptide in such an assay. The Harlil
CC peptides are also useful as a trap material in a diagnostic or
CC therapeutic assay. Therefore, the Harlil peptides are useful in binding
CC assays, protein and antibody purification, therapeutics or diagnostics.
CC In particular, the peptides are also useful for diagnosing Alzheimer's
CC disease, Down's syndrome, neuroectodermal tumours, astrocytomas,
CC glioblastomas, hypoxia, ischaemia or cerebral infarction. The peptides
CC are also useful as targets for drug development for the treatment of
CC these diseases. The present sequence represents a neural thread protein
CC given in the exemplification of the present invention
XX
SQ Sequence 375 AA;

Alignment Scores:

Pred. No.:	5.58e-39	Length:	375
Score:	389.50	Matches:	128
Percent Similarity:	46.2%	Conservative:	40
Best Local Similarity:	35.2%	Mismatches:	88
Query Match:	23.8%	Indels:	108
DB:	5	Gaps:	15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABB81538 (1-375)

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||   ||||||   |||   ||||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
              ||||||   :::   ||||||:::|||      |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
              :::|||   ||||||   ||||||      ||||||||||||||||    ||||||
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
                    |||||||||:::
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                  ||||||:::|||      ||||||                          |||   |
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------- 627
              ||:::                        :::|||:::
Db        128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 -------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
                          :::|||:::   |||   |||           :::::::::
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159
```

```
Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                   :::|||:::   |||         |||
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
                |||   |||                   |||   |||::::::   :::
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                  |||         :::|||   :::      :::   :::|||
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ----------------------------------------TTTGGCTTGTTTACT- 450
                                                     |||   ||||||:::
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ------------------------------------------------------TTTA 446
                                                                     |||:
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
                ::      :::          |||:::         |||      |||:::
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
                :::|||      ||||||              |||:::   ||||||      :::
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
                  ::::::|||   :::      |||||:::||||||      |||   |||||||||
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
                ||||||||   :::|||||||||||||||||||||||||      |||:::||||||
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
                ||||||||||||||||||||||||||||||   |||         |||||||||||||||
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
                ||||||||   |||      ||||||||||||      |||||||||:::|||
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
                |||:::
Db        374 LeuArg 375
```

```
RESULT 11
AAE29142
ID   AAE29142 standard; protein; 375 AA.
XX
AC   AAE29142;
XX
DT   27-JAN-2003  (first entry)
XX
DE   AD7c-NTP protein.
XX
KW   Neural thread protein; NTP; hyperplasia; hypertrophy; arteriosclerosis;
KW   haemorrhoid; gene therapy; tumour; vascular disease; atherosclerosis;
KW   inflammatory disease; nutritional deficiency disease; genetic disease;
KW   autoimmune disease; metabolic disease; traumatic disease; intoxication;
KW   infectious disease; congenital malformation; enzyme deficiency disease;
KW   amyloid disease; fibrosis disease; storage disease; radiation disease;
KW   poisoning; environmental disease; endocrine disease; protein therapy;
KW   degenerative disease; mechanical disease; AD7c-NTP protein.
XX
OS   Unidentified.
XX
PN   WO200274323-A2.
XX
PD   26-SEP-2002.
XX
PF   08-MAR-2002; 2002WO-IB001959.
XX
PR   08-MAR-2001; 2001US-0273957P.
XX
PA   (AVER/) AVERBACK P.
XX
PI   Averback P;
XX
DR   WPI; 2002-759864/82.
DR   N-PSDB; AAD46671.
XX
PT   Treating a condition in a patient requiring removal or destruction of
PT   cells, such as a benign or malignant tumor of a tissue or an inflammatory
PT   disease, comprises administering a neural thread protein (NTP) or a NTP
PT   gene to a mammal.
XX
PS   Example 2; Fig 1; 70pp; English.
XX
CC   The invention relates to a method for treating a condition in a patient
CC   requiring removal or destruction of cells. The method involves
CC   administering to a mammal a neural thread protein (NTP), or administering
CC   to a tumour or other target cell a NTP gene, where the expression of the
CC   NTP gene is induced resulting in expression of the NTP protein. The
CC   method and NTP are useful for treating a condition in a patient requiring
CC   removal or destruction of cells, such as a benign or malignant tumour of
CC   a tissue, a hyperplasia, hypertrophy, or overgrowth of a tissue,
CC   preferably tonsillar hypertrophy or prostatic hyperplasia, a virally,
CC   bacterially, or parasitically altered tissue, or a malformation of a
CC   tissue. Other conditions include a cosmetic modification to a tissue,
CC   such as removal of unwanted facial hair, warts or unwanted fatty tissue,
CC   a vascular disease, particularly atherosclerosis or arteriosclerosis,
CC   haemorrhoids, or varicose veins, an inflammatory disease, autoimmune
CC   disease, metabolic disease, hereditary/genetic disease, traumatic disease
CC   or physical injury, nutritional deficiency disease, infectious disease,
CC   congenital malformation, amyloid disease, fibrosis disease, storage
CC   disease, enzyme deficiency disease, poisoning, intoxication, degenerative
CC   disease, radiation disease, environmental disease, endocrine disease or
CC   mechanical disease. The invention is useful in protein therapy and gene
CC   therapy. The present sequence is AD7c-NTP protein
XX
SQ   Sequence 375 AA;
```

```
Alignment Scores:
Pred. No.:              5.58e-39        Length:         375
Score:                  389.50          Matches:        128
Percent Similarity:     46.2%           Conservative:   40
Best Local Similarity:  35.2%           Mismatches:     88
Query Match:            23.8%           Indels:         108
DB:                     5               Gaps:           15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x AAE29142 (1-375)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
                |||:::|||||||||   |||   |||||||||   ||||||   |||   |||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
                ||||||   :::   ||||||:::|||      |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
                :::|||   ||||||   ||||||      ||||||||||||||||||    ||||||
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
                    |||||||||:::
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                     ||||||:::|||      ||||||                        |||   |
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------- 627
```

```
           ||:::                          :::|||:::
Db        128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 -------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
                          :::|||:::   |||   |||           :::::::::
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                  :::|||:::   |||       |||
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
                 |||   |||                  |||   |||::::::   :::
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                  |||         :::|||   :::       :::   :::|||
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ------------------------------------------TTTGGCTTGTTTACT- 450
                                                        |||   ||||||:::
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ----------------------------------------------------TTTA 446
                                                                   |||:
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
                ::      :::          |||:::         |||       |||:::
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
                ::|||      ||||||          |||:::   ||||||       :::
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
                  ::::::|||   :::      ||||||:::||||||       |||   |||||||||
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
                |||||||||   :::||||||||||||||||||||||||||||      |||:::||||||
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
                |||||||||||||||||||||||||||||    |||       ||||||||||||||
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
                |||||||||   |||      ||||||||||||       |||||||||:::|||
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
                |||:::
Db        374 LeuArg 375
```

RESULT 12
ABR63268
ID ABR63268 standard; protein; 375 AA.
XX
AC ABR63268;
XX
DT 28-AUG-2003 (first entry)
XX
DE AD7c-NTP protein.
XX
KW Cytostatic; Antibacterial; Immunosuppressive; Antiinflammatory;
KW neural thread protein; NTP; tumour.
XX
OS Unidentified.
XX
PN WO2003008443-A2.
XX
PD 30-JAN-2003.
XX
PF 19-JUL-2002; 2002WO-CA001105.
XX
PR 19-JUL-2001; 2001US-0306150P.
PR 19-JUL-2001; 2001US-0306161P.
PR 16-NOV-2001; 2001US-0331477P.
XX
PA (NYMO-) NYMOX CORP.
XX
PI Averback PA;
XX
DR WPI; 2003-247999/24.
XX
PT Novel neural thread protein peptide, referred as cell death peptide,
PT useful for treating prostatic hyperplasia, psoriasis, eczema, dermatosis,
PT atherosclerosis, cosmetic modification to skin, throat, mouth, muscle.
XX
PS Disclosure; Fig 1; 77pp; English.
XX
CC The present invention relates to a neural thread protein (NTP) peptide
CC referred to as cell death peptide. Thought to be cytostatic,
CC antibacterial, immunosuppressive and antiinflammatory. It is useful for
CC treating a condition in a patient requiring removal or destruction of
CC cells, for treating a condition such as benign or malignant tumor,
CC inflammatory disease, autoimmune disease and infectious disease. The
CC peptide useful for treatment is derived from the amino acid sequence for
CC a pancreatic thread protein. The peptide is conjugated, linked or bound
CC to a molecule chosen from antibody or its fragment, antibody-like binding
CC molecule, where the molecule has a higher affinity for binding to a tumor
CC or other target than binding to other cells. Treatment using NTP peptides
CC can remove benign tumors with less risk and fewer of the undesirable side
CC effects of surgery. The present sequence is an NTP amino acid sequence
XX
SQ Sequence 375 AA;

Alignment Scores:

Pred. No.:	5.58e-39	Length:	375
Score:	389.50	Matches:	128
Percent Similarity:	46.2%	Conservative:	40
Best Local Similarity:	35.2%	Mismatches:	88
Query Match:	23.8%	Indels:	108
DB:	6	Gaps:	15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABR63268 (1-375)

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
                |||:::|||||||||   |||   |||||||||   ||||||   |||   ||||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
                ||||||   :::   ||||||:::|||       |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
                :::|||   ||||||   ||||||       |||||||||||||||||    ||||||
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
                          |||||||||:::
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                     ||||||:::|||       ||||||                        |||   |
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------- 627
                ||:::                          :::|||:::
Db        128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 -------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
                          :::|||:::   |||   |||           :::::::::
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159
```

```
Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ----------------------------------------------TTTGGCTTGTTTACT- 450
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ------------------------------------------------------TTTA 446
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
Db        374 LeuArg 375
```

```
RESULT 13
ADA84017
ID   ADA84017 standard; protein; 375 AA.
XX
AC   ADA84017;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human POM80 protein.
XX
KW   human; marker; expressed sequence tag; EST; arabidopsis; tumour;
KW   stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2002103028-A2.
XX
PD   27-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-IB004189.
XX
PR   30-MAY-2001; 2001US-0293999P.
PR   22-OCT-2001; 2001US-0330457P.
PR   19-FEB-2002; 2002US-0357144P.
XX
PA   (BIOM-) BIOMEDICAL CENT.
XX
PI   Baranova AV,  Yankovsky NK,  Kozlov AP,  Lobashev AV,  Krukovskaya LL;
XX
DR   WPI; 2003-175241/17.
DR   N-PSDB; ADA84016.
XX
PT   Determining if a nucleic acid is a marker for a phenotype/cell type of
PT   interest, by global comparison of expressed sequence tags known to be
PT   expressed in the phenotype/cell type with all ESTs expressed in normal
PT   tissue.
XX
PS   Claim 29; Page 398-400; 516pp; English.
XX
CC   The invention relates to a novel method for determining if a nucleic acid
CC   is a marker for a predetermined phenotype/cell type of interest from a
CC   biological species. The method comprises performing a global comparison
CC   of a group of expressed sequence tags (ESTs) known to be expressed in the
CC   phenotype/cell type of interest with all ESTs expressed in normal tissue
CC   in order to identify ESTs that are preferentially expressed in the
CC   phenotype/cell of interest. A method of the invention is useful for
CC   determining whether a nucleic acid is a marker for a predetermined
CC   phenotype or cell type of interest from a biological species, preferably
CC   Arabidopsis or human. The cell type of interest is an abnormal cell such
CC   as a tumour cell, and the predetermined phenotype is a stress-induced
CC   phenotype such as hyperosmotic stress or high salt conditions. A method
CC   of the invention is also useful for determining the progression of colon
CC   cancer in a human, for detecting a tumour cell, and for regulating or
CC   preventing the growth of a tumour cell. An antibody of the invention is
CC   useful for detecting the absence or presence of peptides encoded by
CC   tumour-associated markers. A polypeptide of the invention is useful as an
CC   immunogen for vaccinating an animal. The present sequence represents a
CC   tumour-associated antigen of the invention.
XX
SQ   Sequence 375 AA;
```

```
Alignment Scores:
Pred. No.:                5.58e-39      Length:          375
Score:                    389.50        Matches:         128
Percent Similarity:       46.2%         Conservative:    40
Best Local Similarity:    35.2%         Mismatches:      88
Query Match:              23.8%         Indels:          108
DB:                       6             Gaps:            15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ADA84017 (1-375)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT------------GTATTTTAAAAATTAT 668
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------- 627
Db        128 rpAsp--------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 --------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
```

```
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                  :::|||:::   |||         |||
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
                  |||   |||                 |||   |||:::::   :::
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                  |||         :::|||   :::      :::   :::|||
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ------------------------------------------TTTGGCTTGTTTACT- 450
                                                        |||   ||||||:::
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ----------------------------------------------------TTTA 446
                                                                     |||:
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
                  ::      :::      |||:::      |||      |||:::
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
                ::|||      ||||||               |||:::   ||||||      :::
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
                  ::::::|||   :::      ||||||:::||||||      |||   |||||||||
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
              |||||||||   :::|||||||||||||||||||||||||||      |||:::||||||
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
              |||||||||||||||||||||||||||||   |||         |||||||||||||||
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
              |||||||||   |||      ||||||||||      |||||||||:::|||
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
              |||:::
Db        374 LeuArg 375
```

RESULT 14
ABU03024
ID ABU03024 standard; protein; 375 AA.
XX
AC ABU03024;
XX
DT 20-JAN-2003 (first entry)
XX
DE Human neural thread protein AD7C-NTP.
XX
KW Neural thread protein; NTP-peptide; AD7C-NTP; surgical excision;
KW transplantation; grafting; chemotherapy; immunotherapy; vaccination;
KW ablation; cryotherapy; laser therapy; phototherapy; gene therapy;
KW radiation; tumour; hyperplasia; hypertrophy; overgrowth of tissue;
KW malformation of tissue; tonsillary hypertrophy; prostatic hyperplasia;
KW cosmetic modification; vascular disease; atherosclerosis;
KW arteriosclerosis; haemorrhoid; varicose vein; inflammatory disease;
KW autoimmune disease; metabolic disease; traumatic disease;
KW physical injury; nutritional deficiency disease; infectious disease;
KW amyloid disease; fibrosis disease; storage disease;
KW congenital malformation; enzyme deficiency disease; poisoning;
KW intoxication; environmental disease; radiation disease;
KW endocrine disease; degenerative disease; mechanical disease.
XX
OS Homo sapiens.
XX
PN WO200297030-A2.
XX
PD 05-DEC-2002.
XX
PF 24-MAY-2002; 2002WO-CA000759.
XX
PR 25-MAY-2001; 2001US-0293156P.
XX
PA (NYMO-) NYMOX CORP.
XX
PI Averback PA;
XX
DR WPI; 2003-041406/03.
XX
PT Novel peptides similar in amino acid sequence to neural thread proteins
PT (NTP), useful for treating unwanted cellular proliferations such as
PT malignant tumors and prostatic hyperplasia.
XX
PS Disclosure; Fig 1; 78pp; English.
XX
CC The invention describes an NTP-peptide (I) comprising at least one amino
CC acid sequence corresponding to part of the amino acid sequence of a
CC neural thread protein, AD7C-NTP. The invention provides a method of
CC treating a condition requiring removal or destruction of cells of a
CC mammal comprising administering to a mammal, a therapeutic amount of (I).
CC The treatment is administered to the mammal before, during or after
CC surgical excision, transplantation, grafting, chemotherapy,
CC immunotherapy, vaccination, thermal or electrical ablation, cryotherapy,
CC laser therapy, phototherapy, gene therapy and/or radiation. The method is
CC useful for treatment of benign or malignant tumour; hyperplasia,
CC hypertrophy or overgrowth of tissue; virally, bacterially or
CC parasitically altered tissue; malformation of tissue selected from lung,
CC breast, stomach, pancreas, prostate, bladder, bone, ovary, skin, kidney,
CC sinus, colon, intestine, rectum, esophagus, heart, spleen, salivary
CC gland, blood, brain and its coverings, spinal cord, muscle, connective
CC tissue, adrenal, parathyroid, thyroid, uterus, testis, pituitary,
CC reproductive organs, liver, hair, gall bladder, eye, ear, nose, throat,
CC tonsils, mouth and lymph nodes and lymphoid system; tonsillary
CC hypertrophy; prostatic hyperplasia; cosmetic modification to a tissue;
CC vascular disease (atherosclerosis or arteriosclerosis); haemorrhoids;
CC varicose veins; inflammatory disease; autoimmune disease; metabolic
CC disease; hereditary/genetic disease; traumatic disease; physical injury;
CC nutritional deficiency disease; infectious disease; amyloid disease;
CC fibrosis disease; storage disease; congenital malformation; enzyme
CC deficiency disease; poisoning; intoxication; environmental disease;
CC radiation disease; endocrine disease; degenerative disease and mechanical
CC disease. This is the amino acid sequence of the human neural thread
CC protein AD7C-NTP
XX
SQ Sequence 375 AA;

Alignment Scores:

Pred. No.:	5.58e-39	Length:	375
Score:	389.50	Matches:	128
Percent Similarity:	46.2%	Conservative:	40
Best Local Similarity:	35.2%	Mismatches:	88
Query Match:	23.8%	Indels:	108
DB:	6	Gaps:	15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABU03024 (1-375)

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||  ||||||   |||   ||||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
              ||||||   :::   ||||||:::|||      |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
```

```
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------ 627
Db        128 rpAsp--------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 --------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ----------------------------------------TTTGGCTTGTTTACT- 450
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ------------------------------------------------------TTTA 446
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy         94 CTCAAG 89
Db        374 LeuArg 375

RESULT 15
ABB99774
ID   ABB99774 standard; protein; 375 AA.
XX
AC   ABB99774;
XX
DT   24-MAR-2003  (first entry)
XX
DE   Amino acid sequence of human neuronal thread protein AD7c-NTP.
XX
KW   Human; neuronal thread protein; AD7c-NTP; Alzheimer's disease; histone;
KW   neurodegeneration; in vivo gene expression; amphipathic compound;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200299036-A2.
XX
PD   12-DEC-2002.
XX
PF   28-MAY-2002; 2002WO-US016429.
XX
PR   01-JUN-2001; 2001US-00872968.
XX
PA   (RHOD-) RHODE ISLAND HOSPITAL.
XX
PI   Wands JR,  De La Monte SM;
XX
DR   WPI; 2003-140605/13.
DR   N-PSDB; ABZ23236.
XX
PT   Inducing prolonged in vivo gene expression in mammal by contacting
PT   neuronal tissue with composition comprising Alzheimer's disease-
PT   associated neural thread protein 7c antisense nucleic acid, histone,
PT   amphipathic compound.
XX
PS   Disclosure; Page 35; 69pp; English.
XX
CC   The present sequence represents a human neuronal thread protein AD7c-NTP.
CC   AD7c-NTP is overexpressed in brains with Alzheimer's disease at early and
CC   intermediate stages of neurodegeneration. The expression of AD7c-NTP may
CC   be reduced using the method of the invention. The specification describes
CC   a method for inducing prolonged in vivo gene expression in a mammal. The
CC   method comprises contacting a non-muscular tissue with a composition
CC   comprising a nucleic acid, histone and an amphipathic compound. The
CC   method is useful for inducing prolonged in vivo gene expression in non-
CC   muscular tissue of a mammal, e.g. neuronal tissue, central nervous system
CC   (CNS) tissue, tissue comprising a post-mitotic neuronal cell, cortical
CC   neuronal cell or hippocampal neuronal cell, glial cell, or vascular
CC   endothelial cell. The method is useful in gene therapy applications to
CC   treat Alzheimer's disease , where the composition comprises antisense
CC   AD7c-NTP nucleic acid
XX
SQ   Sequence 375 AA;

Alignment Scores:
Pred. No.:              5.58e-39        Length:         375
Score:                  389.50          Matches:        128
Percent Similarity:     46.2%           Conservative:   40
Best Local Similarity:  35.2%           Mismatches:     88
Query Match:            23.8%           Indels:         108
DB:                     6               Gaps:           15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ABB99774 (1-375)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
```

```
Db        88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy       707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                 ||||||:::|||      ||||||                        |||   |
Db       108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy       667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------ 627
                ||:::                      :::|||:::
Db       128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy       626 -------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
                         :::|||:::   |||   |||          :::::::::
Db       139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy       587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                   :::|||:::   |||          |||
Db       159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy       556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
                 |||   |||                  |||   |||::::::   :::
Db       179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy       514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                |||         :::|||   :::      :::   :::|||
Db       197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy       464 ----------------------------------------TTTGGCTTGTTTACT- 450
                                                    |||   ||||||:::
Db       214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy       449 ------------------------------------------------------TTTA 446
                                                                      |||:
Db       234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy       445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
                ::      :::      |||:::     |||      |||:::
Db       254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy       391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
               :::|||      ||||||               |||:::   ||||||      :::
Db       274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy       331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
                ::::::|||   :::      ||||||:::||||||.     |||   |||||||||
Db       294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy       274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
                |||||||||   :::|||||||||||||||||||||||||||      |||:::||||||
Db       314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy       214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
                ||||||||||||||||||||||||||||||   |||         |||||||||||||||
Db       334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy       154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
                |||||||||   |||      ||||||||||||      |||||||||:::|||
Db       354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373

Qy        94 CTCAAG 89
                |||:::
Db       374 LeuArg 375


Search completed: February  8, 2006, 10:42:44
Job time : 135.475 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07 ; Search time 6.14867 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 1640
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092020_27698/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_92_@runat_08022006_092020_27698 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	388	23.7	613	4	C40201	artifact-warning s
c	2	357.5	21.8	673	4	F40201	artifact-warning s
c	3	352	21.5	627	4	A40201	artifact-warning s
	4	350	21.3	673	4	F40201	artifact-warning s
	5	335	20.4	627	4	A40201	artifact-warning s
	6	327	19.9	613	4	C40201	artifact-warning s
	7	318.5	19.4	418	2	S41044	chromosomal protei
c	8	286	17.4	597	4	E40201	artifact-warning s
	9	264	16.1	196	2	I38022	hypothetical prote
	10	256.5	15.6	597	4	E40201	artifact-warning s
c	11	232.5	14.2	100	2	A46010	X-linked retinopat
c	12	214.5	13.1	499	2	S65657	alpha-1C-adrenergi
c	13	182	11.1	579	4	D40201	artifact-warning s
	14	173.5	10.6	579	4	D40201	artifact-warning s
c	15	172	10.5	46	2	I54375	gene NF2 protein -
c	16	171	10.4	301	4	B40201	artifact-warning s
	17	166	10.1	447	2	A57034	transcription fact
c	18	162	9.9	39	2	I54374	gene NF2 protein -
c	19	160	9.8	369	2	A53959	thromboxane A-2 re
c	20	156	9.5	53	2	A42442	integrin beta-1 ch
c	21	155.5	9.5	407	2	T02670	probable thromboxa
c	22	154.5	9.4	79	2	A56194	thromboxane A-2 re
	23	152	9.3	301	4	B40201	artifact-warning s
	24	148.5	9.1	114	2	JC5238	galactosylceramide
c	25	147	9.0	841	1	I78885	serine/threonine-s
c	26	115.5	7.0	619	2	A60646	transforming prote
	27	113.5	6.9	124	2	A47582	B-cell growth fact
c	28	109.5	6.7	440	2	A26359	decay-accelerating
c	29	108.5	6.6	331	2	A54295	interferon alpha/b
c	30	108.5	6.6	331	2	S59501	interferon recepto
c	31	101.5	6.2	107	2	C48078	biliary glycoprote
c	32	91	5.5	252	2	T48681	hypothetical prote
c	33	91	5.5	522	2	T08711	gamma-adaptin homo
c	34	90	5.5	339	2	S09264	Ig alpha chain C r
c	35	89	5.4	347	2	S09274	Ig alpha chain C r
c	36	83	5.1	352	2	S09266	Ig alpha chain C r
c	37	81.5	5.0	1559	1	S64757	probable membrane
c	38	80.5	4.9	350	2	T15178	hypothetical prote
	39	80.5	4.9	1231	2	T24415	hypothetical prote
	40	80	4.9	325	2	T26180	hypothetical prote
c	41	80	4.9	342	2	PC4211	hepatocellular car
c	42	79	4.8	101	2	S60434	probable membrane
c	43	79	4.8	348	2	S09270	Ig alpha chain C r
c	44	78	4.8	357	2	S09267	Ig alpha chain C r
c	45	78	4.8	1148	2	A71446	hypothetical prote

ALIGNMENTS

RESULT 1
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

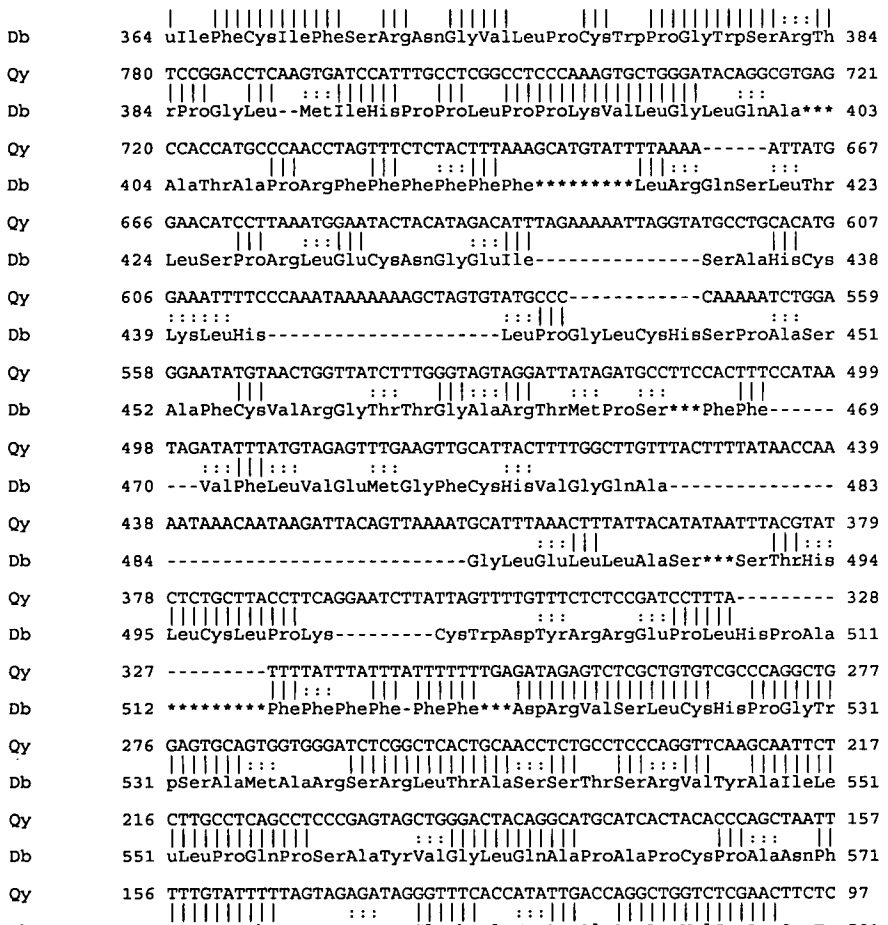
Alignment Scores:

Pred. No.:	2.6e-35	Length:	613
Score:	388.00	Matches:	119
Percent Similarity:	50.8%	Conservative:	36
Best Local Similarity:	39.0%	Mismatches:	100
Query Match:	23.7%	Indels:	51
DB:	4	Gaps:	8

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x C40201 (1-613)

```
Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
              :::|||||||||   ||||||   :::||||||||||||       ||||||||||||||
Db        345 PheSerCysLeuSerLeuLeuArgThrTrpAspTyrArg-ArgProHisHisAlaGlnLe 364

Qy        839 AAT-TTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAAC 781
```



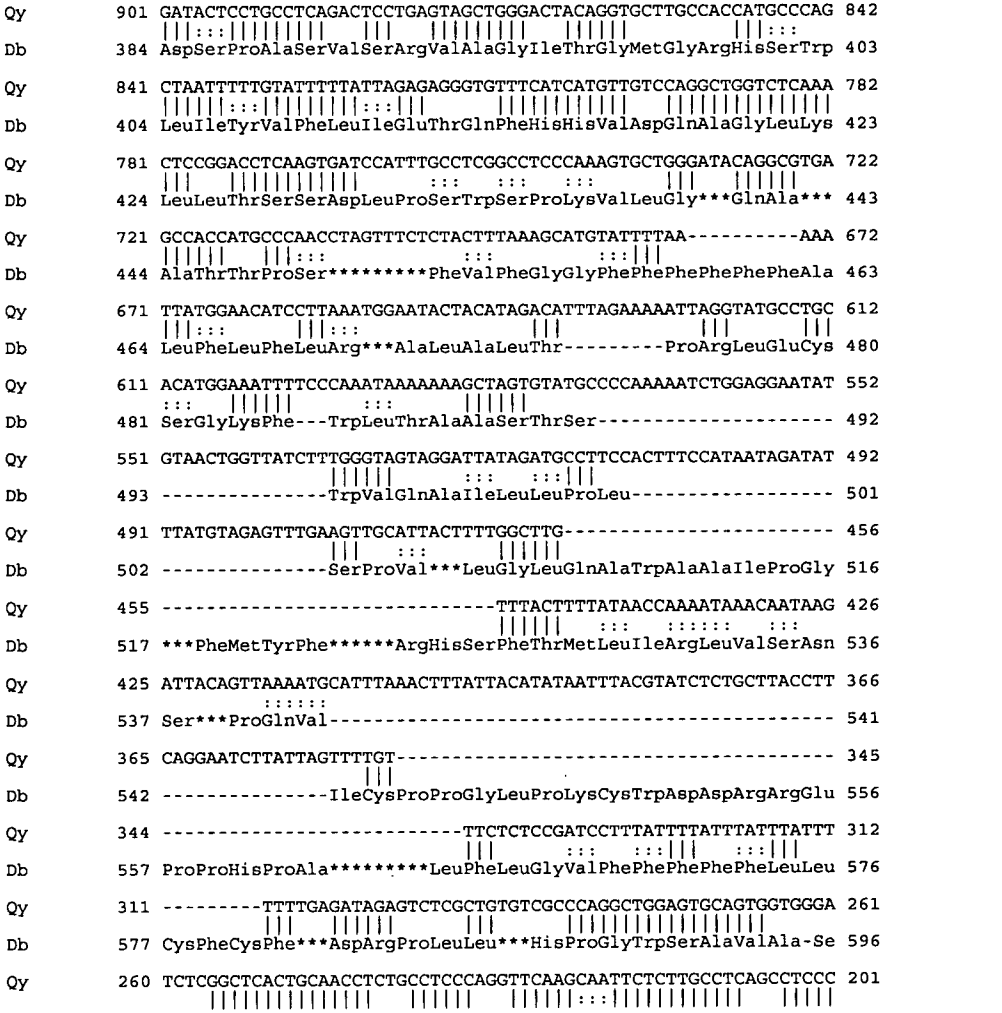
RESULT 2
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.

Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	7.63e-32	Length:	673
Score:	357.50	Matches:	115
Percent Similarity:	43.1%	Conservative:	31
Best Local Similarity:	33.9%	Mismatches:	98
Query Match:	21.8%	Indels:	96
DB:	4	Gaps:	9

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x F40201 (1-673)



```
Db        596 rPheGlySerLeuGlnProGlnProProGlyPheLysArgPheSerCysLeuCysLeuPr 616

Qy        200 GAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAG 141

Db        616 oCysSerTrpAspTyrArgHisGlyProProPheLeuAlaAsnLeuCysIlePheAsnAr 636

Qy        140 AGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCAC 81

Db        636 gAspThrValSerProCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAl 656

Qy         80 CCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCACTGTGCCCAA 28

Db        656 a-LeuLeuValSerGlnSerAlaGlyMetThrGlyValSerHisHisThrGln 673

RESULT 3
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              3.2e-31       Length:          627
Score:                  352.00        Matches:         117
Percent Similarity:     50.2%         Conservative:     36
Best Local Similarity:  38.4%         Mismatches:       99
Query Match:            21.5%         Indels:           53
DB:                     4             Gaps:             11

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x A40201 (1-627)

Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840

Db        355 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProHisTyrAlaArgLeu 374

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780

Db        375 IlePheCysIlePheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSer 394

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGAGC 720

Db        395 ProAspLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGlnAla***A 414

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATC 660

Db        414 spArgAlaArg***---******PhePheIlePheCys---------------------- 425

Qy        659 CTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTT 600

Db        426 ----------PheTyr-----------------------------------------PheP 429

Qy        599 TCCCAAATAAAAAAAGCTAGTGTATGCCC------CAAAAATCTGGAGGAATA------- 553

Db        429 heProArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSerAlaH 449

Qy        552 --TGTAACTGGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGA 495

Db        449 isCysLysLeuArgLeuProGlySerArgHisSerProAla-SerAla------SerGln 466

Qy        494 TATTTATGTAGAGTTTGAAGTTGCATTACTTTTGGC---TTGTTTACTTTTATAACCAAA 438

Db        467 ValAlaGlyThrThrGlyAlaArgThrThrProGly***PhePheValPheLeuValGlu 486

Qy        437 ATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA---TATAATTTACGT 381

Db        487 ThrGlyPheHisArgGlySerGlnAspGlyLeuAspLeuLeuThrSer***SerAlaArg 506

Qy        380 ATCTCTGCT--------------TACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCG 336

Db        507 LeuGlyLeuProLysCysTrpAspTyrArgArgGluThr--------------AlaPro 521

Qy        335 ------------ATCCTTTATTTTATTTATTTATTTTTTGAGATAGAGTCTCGCTGTGT 288

Db        522 Gly*********PheLeuPhePheValPheIlePhePheArgAspGlyValSerLeuCys 541

Qy        287 CGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTT 228

Db        542 ArgGlnGlyTrpSerAlaValAlaArgSerArgLeuThrAlaSerSerAlaSerArgVal 561

Qy        227 CAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACA 168

Db        562 HisAlaIleLeuLeuProGlnProProLys***LeuGlyLeuGlnAlaProAlaLeuArg 581

Qy        167 CCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTC 108

Db        582 ProAlaAsnPheLeuTyrPhe******ArgArgGlyPheThrValValAlaArgMetVal 601

Qy        107 TCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAG 48

Db        602 SerIleSer***ProArg--AspProProAlaLeuAlaSerGlnSerAlaGlyIleThrG 621

Qy         47 GCTTG 43

Db        621 lyVal 622

RESULT 4
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:
Pred. No.:              5.43e-31      Length:          673
Score:                  350.00        Matches:         116
Percent Similarity:     47.0%         Conservative:     33
Best Local Similarity:  36.6%         Mismatches:      114
Query Match:            21.3%         Indels:           54
DB:                     4             Gaps:              6

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x F40201 (1-673)

Qy         29 TGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCA 88
```

```
Db          2 TrpValTrpTrpLeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHi 21

Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
              |||||||      |||   |||||||||   |||:::||||||||:::|||||||||||
Db         21 sLeuArgSerGlyVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLy 41

Qy        149 AATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCT 208
              ||||   |||||||||                 ||||||   |||||||||   ||||||
Db         41 sIleHisLysLeuAlaArgAsnGlyGlyProCysLeu***SerGlnLeuHisGlyArgGl 61

Qy        209 GAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCA 268
              ||||||||||||   ||||||   ||||||   |||||||||||||||:::       |
Db         61 nArgGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGluProLysLeu-AlaT 81

Qy        269 CTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT 328
              ||||||||||||||||      |||:::   ||||||||||||   |||      |||
Db         81 hrAlaLeuGlnProGlyCysGlnSerLysGlyLeuSerGlnLysGlnLysGlnSerLys- 100

Qy        329 AAAGGATCGGAGAGAAACAAAACTAATAAG---------------------------- 358
              |||      ::::::          ||||||
Db        101 LysLysLysLysLysThrProLysAsnLys*********AlaGlyCysGlyGlySerArg 120

Qy        359 ------------------------------------ATTCCTGAAGGTAAGCAGAGATAC 382
                                                   |||      |||::::::
Db        121 LeuSerSerGlnHisPheGlyArgProGlyGlyGlnIleThr***GlyGlnGluPheGlu 140

Qy        383 GTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGG 442
                 ::: :::   |||   |||                       |||   :::|||
Db        141 ThrSer-LeuIleAsnMetVal------------------LysLeuCysLeuTyr***Ly 154

Qy        443 TTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTAT 502
              |||   :::          :::            |||   :::
Db        154 sTyrIleAsn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyrThrGly----- 172

Qy        503 GGAAAGTGGAAGGCATCTA---------TAATCCTACTACCCAAAGATAACCAGTTACAT 553
               :::   ||||||   :::                          ::::::      ::
Db        173 -AspArgGlyArgArgIleAla***ThrGlnGluValGluAlaAlaValSerGlnAsnLe 192

Qy        554 ATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGC 613
              :   |||                        :::      ||||||
Db        192 uProLeuHisSerSerArgGlyValArgAlaArgAlaTyrLeuLysAsn----------- 208

Qy        614 AGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATT 673
                  |||   :::|||                        |||   |||
Db        209 -------LysAsnLysAlaLysLysLysLysLysProProLys-------------- 221

Qy        674 TTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTAT 733
                              ::::::          ||||||:::|||||||||||||||
Db        222 ------------------ThrLys*********LeuGlyValValAlaHisAlaCysHi 235

Qy        734 CCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGC 793
              |||:::            :::         ||||||||||||||||||||||||||||
Db        235 sProSerThrLeuGlyAspGlnGluGlyArgSerLeuGluValArgSerLeuArgProAl 255

Qy        794 CTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTG 853
              |   :::|||      |||   :::|||   |||||||||   ||||||      |||
Db        255 a***SerThrTrp***AsnCysValSerIleLysAsnThr***IleSerGlnGluTrpAr 275

Qy        854 GCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTGAGGCAGGA 896
                  ||||||:::|||||||||:::|||:::|||||||||
Db        275 gProMetProValIleProAlaThrArgGluThrGluAlaGly 289

RESULT 5
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.:                2.73e-29        Length:          627
Score:                    335.00          Matches:         116
Percent Similarity:       43.7%           Conservative:    27
Best Local Similarity:    35.5%           Mismatches:      87
Query Match:              20.4%           Indels:          97
DB:                       4               Gaps:            8

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x A40201 (1-627)

Qy         40 ACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAG 99
                 :::   |||||||||   |||||||||      |||||||||   ||||||  |||
Db          5 SerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyGlyArgIleThr--ArgSerG 24

Qy        100 AAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAAT 159
                    ||||||   ||||||   :::|||||||||   ||||||||||||||||||||:
Db         24 lyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleGlnLysI 44

Qy        160 TAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGA 219
                    :::::                |||   ||||||||||||||||||||||||||||||
Db         44 leSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluA 64

Qy        220 ATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAG 279
              ||   :::|||   ||||||      ||||||||||||||||||   ||||||||||||
Db         64 snGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCysThrProA 84

Qy        280 CCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGA 339
              ||   ||||||:::||||||||||:::|||:::|||      ||||||   |||     :
Db         84 laLeuAlaThrGluArgAspSerValSerGluLys------AsnLysAsnLysLys***L 102

Qy        340 GAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATA 399
              ::                   |||   |||
Db        102 ys*********AlaGlyArgGlyLeu---------------------------------- 110

Qy        400 AAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAG 459
                                         ::::::         :::::::::|||
Db        111 ----------------------ThrProValIleProAlaLeuTrpGluAlaLysAlaG 123

Qy        460 CCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATC 519
                 |||                  |||   ::::::              |||   |
Db        123 lyGlySerArgGlyGlnGluIleGluThrIleLeuAlaThrThrValLysProArg---L 142

Qy        520 TATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACA 579
                 ||          |||   |||:::
Db        142 euTyr***LysTyrLysLysLeuAla---------------------------------- 150

Qy        580 CTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTC 639
                              |||:::                          |||
Db        151 -----------------GlyArg--------------------------SerAlaGlyA 156

Qy        640 TATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTA-----------AAATACATGC 687
                |||||                  |||   :::|||  .          :::   |||
Db        156 laCysSer-----------------ProSerTyrLeuGlyGly***GlyArgArgMetA 170

Qy        688 TTTAAAGTAGAGAAACTAGGTTG---------------------------------- 710
                  :::||||||      |||
Db        170 la***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuGlnProT 190

Qy        710 ---------------------------------------------------------- 710

Db        190 rpArgGlnSerGluThrProSerArgLysLysIleLysThrLysAsnLysLys******* 210
```

```
Qy        711 -----GGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGAT 765
                   |||   |||:::          |||||||||||||||||||||||:::   |||
Db        210 **ProGlyAlaValSerArgLeu-***SerGlnHisPheGlyArgProArgArgAlaAsp 229

Qy        766 CACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATA 825
              |||  |||||||||      ||||||:::|||          |||   :::|||   |
Db        230 His--GluValArgArgSerArgProSerTrpLeuProArg***AsnProValSerThrL 249

Qy        826 AAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGT 885
              ||||||||||                 :::|||||||||||||||||||||   |||:
Db        249 ysAsnThrLysAsn***ProGlyValValArgAlaProValValProAlaThrTrpGluA 269

Qy        886 CTGAGGCAGGA 896
              ::|||||||||
Db        269 laGluAlaGly 272
```

RESULT 6
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	2.2e-28	Length:	613
Score:	327.00	Matches:	116
Percent Similarity:	48.9%	Conservative:	34
Best Local Similarity:	37.8%	Mismatches:	101
Query Match:	19.9%	Indels:	57
DB:	4	Gaps:	9

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x C40201 (1-613)

```
Qy         35 CAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAG 94
              |||   |||   ||||||          ||||||:::::|||||||||||||||     |
Db          4 GlnTrpLeuThrProValIleProAlaLeuTrpGluAlaLysValGlyGlySer***--G 23

Qy         95 GTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACA 154
              ||::::::|||||||||||||||   ||||||   ||||||:::   |||   |||||||
Db         23 lyGlnGluPheGluThrSerLeuAlaAsnMetAlaLysProHisPheTyr***LysTyrL 43

Qy        155 AAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCA 214
              ||          |||             ||||||||||||:::|||:::   |||   |||:
Db         43 ysLysLeuAlaGlyHisGlyAlaGlyAlaCysSerProThrTyrAlaGluGly***GlyA 63

Qy        215 AGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCAC 274
              ::|||:::|||   |||   |||   |||:::|||||||||||||||          ||||
Db         63 rgArgMetAla***ThrArgGluValGluLeuAlaValSerArgAspLeuAlaIleAlaL 83

Qy        275 TCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAAT------ 328
              ||||||||||      ||||||||||||||||||||   ||| |||
Db         83 euGlnProGly***GlnSerGluThrLeuSerGlnLysLysLys-LysLysLys****** 102

Qy        329 ---AAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTA 385
                  |||          ::::::          |||:::||||||:::|||
Db        103 ***AlaGlyCysSerGlySerArgLeu***SerGlnHisPheGlyArgGlnArgTrpVal 122

Qy        386 AATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTA 445
              .. ::::::          |||                                  |||
Db        123 AspHisGluAlaArgSerSerArgPro----------------------AlaTrpPro 134

Qy        446 TAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGA 505
                    |||   |||          :::   |||        |||          |||
Db        135 ---------ThrTrpGlnAsnProIleSerThrLysAsnThrLysAsn---***LeuGly 150

Qy        506 AAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATT--------- 556
                  :::|||   ::::::      |||:::      :::
Db        151 MetValArgAlaProValVal------ProArgThrGlnLysAlaGluAlaGlyGluTrp 168

Qy        557 ---CCTCCAGATTTTTGGGGC----------------------ATACACTAGCTTTTT 589
                  |||      |||                          :::|||
Db        169 HisLysProGlyArgTrpSerLeuGln***AlaGluIleSerProLeuHis---SerSer 187

Qy        590 TTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTAT 649
              |||      ::::::   :::|||
Db        188 LeuGlyAspArgValArgLeuCys---------------------------------- 195

Qy        650 TCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTT 709
                                          |||             ||| :::
Db        196 ----------------------LeuLysLysLysLysLysLys-Lys*********Ar 207

Qy        710 GGGCATGGTGGCTCACGCCTGTAT-CCCAGCACTTTGGGAGGCCGAGGCAAATGGATCAC 768
              |||   ||||||||||||||   ||||||||||||||||:::|||   ||||||
Db        207 gGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyLysGlyGlyTrpIle-- 226

Qy        769 TTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAA 828
              :::|||   ||||||   ||||||||||||||||   :::||||||   |||:::|||
Db        227 MetArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeuLys 246

Qy        829 ATACAAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGAGTCTG 888
              |||||||||::::::                  |||   |||   :::   |||   |||
Db        247 IleGlnLysIleSerTrpAlaTrpCysGlyArgLeu***SerHisValArgArgArgLeu 266

Qy        889 AGGCAGGAG 897
              |||||||||
Db        267 ArgGlnGlu 269
```

RESULT 7
S41044
chromosomal protein - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
C;Accession: S41044
R;Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A;Title: A new chromosomal protein essential for mitotic spindle assembly.
A;Reference number: S41044; MUID:94166884; PMID:8121495
A;Accession: S41044
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-418 <YEO>
A;Cross-references: UNIPROT:P49646; UNIPARC:UPI000013C29A; GB:L26953; NID:g537529; PIDN
C;Superfamily: very very hypothetical protein RMSA-1
C;Keywords: chromosomal protein

Alignment Scores:

Pred. No.:	1.97e-27	Length:	418
Score:	318.50	Matches:	109
Percent Similarity:	44.6%	Conservative:	31
Best Local Similarity:	34.7%	Mismatches:	68
Query Match:	19.4%	Indels:	106
DB:	2	Gaps:	13

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x S41044 (1-418)

```
Qy         59 GTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGG 118
              ::::::|||::::::|||   |||   |||||||||||||||||||||||||||||||
Db        173 LeuIleTrpGluAlaLysAlaGlyGluSerLeuGluValArgSerSerArgProAlaSer 192
```

```
Qy        119 TCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGC 178
Db        193 GlnSerArgArgAsnSerValSerThrLysAsnIleLysIleSer--------------- 207

Qy        179 ATGCCTGTA--------------------------------------------------- 187
Db        208 ---ProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGlyVal 226

Qy        188 ---------GTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGG 238
Db        227 HisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAspPro-GlyGl 246

Qy        239 CAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGA 298
Db        246 yArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThrSer---Gl 265

Qy        299 CTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAG 358
Db        265 uThrValSerLysLys------LysLysLysLysLysLysLysThrArgValLeuTh 283

Qy        359 ATTCCTGAAGGTAAGCAGA-----------GATACGTAAATTATATGTAATAAA----- 401
Db        283 rCysIleAsnAlaSerThrLeuPheHisValLeuThrArgIlePheCysTyrLysGlnLy 303

Qy        402 ---------------GTTTAAATGCATTTTAACTGTAATCTTATTGTTTAT-------- 437
Db        303 sTyrProLeuTrpLeuLeuIleLysArgLeuSerValAsnLeuGlnLysGlnGlnGluAs 323

Qy        438 -TTTGGTTATAAAAGTAAACAAGCCAAAAGT---------AATGCAACTTCAAACTCTAC 487
Db        323 pLeuGlnArgLysLysLysGlnArgLysGlnGluHisProGluLeuThrAlaGlnAsnTh 343

Qy        488 ATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAG 547
Db        343 rValLeuLysMetLeuLeuGlnGly---ValLeuIleAsnLeuTyrProThrValThrLe 362

Qy        548 TTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCC 607
Db        362 u------------------------------------------------------------ 362

Qy        608 ATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCC 667
Db        362 ------------------------------------------------------------ 362

Qy        668 ATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGC 727
Db        363 ---------------------------------AsnArgProGlyValValAlaHisAl 371

Qy        728 CTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAG 787
Db        371 aCysIleProAlaLeuTrp---AlaGluGlyGly-ValHisLeuAsnProGly------- 387

Qy        788 ACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGC 847
Db        388 --GlnProGlyGlnHisGlyGluThrProCysLeuGlnLysIleLysLys---Pro-Gly 405

Qy        848 ATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGA 883
Db        406 ValValAlaSerSerTyrSerSerSerTyrLeuGly 417

RESULT 8
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.:              9.98e-24        Length:         597
Score:                  286.00          Matches:        116
Percent Similarity:     45.0%           Conservative:   25
Best Local Similarity:  37.1%           Mismatches:     74
Query Match:            17.4%           Indels:         102
DB:                     4               Gaps:           10

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x E40201 (1-597)

Qy        898 ACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCTA 839
Db        334 SerProValSerAlaSerArgValAlaGlyIleThrGlyMetArgHisHisAlaGlnLeu 353

Qy        838 ATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACTC 779
Db        354 Ile-LeuTyrPhe******ArgTrpGlyPheSerMetLeuValLysLeuValSerAsnSe 373

Qy        778 CGGACCTCAAGTGATCCATTT-GCCTCGGCCTCCCAAAGTGCTGGGAT-ACAGGCGTGAG 721
Db        373 rGlnProGlnValIleArgProAlaLeuAsnSerGlnSerAlaGlyIleThrGlyMetSe 393

Qy        720 CCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACAT 661
Db        393 rTyrHis------------------------------------------------------ 395

Qy        660 CCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAA-- 603
Db        396 ----------------------------------------------------ThrTrp***** 399

Qy        602 ----------TTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATA 553
Db        399 ****PhePhePhePheGluThrGluPheCysSerCysCysProGlyArg----------- 415

Qy        552 TGTAACTGGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTC------CATAA 499
Db        416 ----------------------------------------SerThrMetAlaGlnSerHi 422

Qy        498 TAGATATTTATGTAGAGTTTGAAGT---------------------------------- 474
Db        422 sArgLysLeuCysPheLeuGlySerSerAsnSerLeuLeuSerGlnProProGlu***Le 442

Qy        473 -------------TGCATTACT-----------TTTGGCTTGTTTACTTTTATAACCAA 439
Db        442 uGlyLeuGlnAlaCysAlaThrThrProSer***PheTyrIlePheSer----------- 458

Qy        438 AATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTAT 379
Db        459 ----ArgAspGlyValSerProCysTrpSerSerTrpSerGlnThrProAsnLeuArg** 477

Qy        378 CTCTGCTTA----------CCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTT 329
Db        477 *SerAlaProPro***ThrProLysAlaLeuGlyLeuGlnAla***AlaThrThrProGl 497

Qy        328 ATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGT-GTCGCCCAGGCTGGAGTGCA 270
Db        497 y*********PhePhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValAr 517

Qy        269 GTGGTGGGATCTCGGCTCA--------------CTGCAACCTCTGCCTCCCAGGTTCAA 225
Db        517 gTrpHisAsnLeuThrAlaAsnPheAlaSerTrpValGlnAlaIle-------------- 532

Qy        224 GCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCC 165
```

```
Db    533 ----LeuSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArgPr 551

Qy    164 AGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCG 105
Db    551 oAlaAsn-PheIlePheLeuValGluMetGlyPheLeuHisValGlyGlnAlaGlyLeuL 571

Qy    104 AACTTCTCACCTCAAGTGATCCACCCACCTTGGAC 70
Db    571 ysLeuProThrSerGlyAspProProArgLeuGlu 582
```

RESULT 9
I38022
hypothetical protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C;Accession: I38022
R;Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A;Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A;Reference number: I38021
A;Accession: I38022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI

Alignment Scores:

Pred. No.:	2.85e-21	Length:	196
Score:	264.00	Matches:	61
Percent Similarity:	68.2%	Conservative:	12
Best Local Similarity:	57.0%	Mismatches:	23
Query Match:	16.1%	Indels:	12
DB:	2	Gaps:	3

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x I38022 (1-196)

```
Qy     87 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 146
Db      3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22

Qy    147 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 206
Db     23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42

Qy    207 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 266
Db     43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62

Qy    267 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 326
Db     63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGln------AsnLys 80

Qy    327 ATAAAGGATCGGAG-------AGAAACAAAACTAATAAGATT------------------ 361
Db     81 ThrLysAsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100

Qy    362 ---CCTGAAGGTAAGCAGAGA 379
Db    101 GlyProGluGlyGluGluLys 107
```

RESULT 10
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:

Pred. No.:	2.24e-20	Length:	597
Score:	256.50	Matches:	112
Percent Similarity:	45.5%	Conservative:	26
Best Local Similarity:	37.0%	Mismatches:	99
Query Match:	15.6%	Indels:	67
DB:	4	Gaps:	10

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x E40201 (1-597)

```
Qy     41 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG-ATCACTTGAGGTGAG 99
Db      6 LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArgIleThr***GlyTrp 25

Qy    100 AAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAAT 159
Db     26 GluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr***LysTyr-LysIl 45

Qy    160 TAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG---GCAAG 216
Db     45 eSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGluThrGly-G 65

Qy    217 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 276
Db     65 luAsnCysLeuAsnProGlySerLysValCysGlyGlu-IleValProSerTyrSer--- 83

Qy    277 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATC 336
Db     84 GlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys*********AlaArgCys 103

Qy    337 GGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTA 396
Db    104 GlySer------------------------------------------------------ 105

Qy    397 ATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTGGTTATAAAAGTAAAC 456
Db    106 ---------SerCysLeu***Ser------GlnArgPheGlySerSerArgArgGlyGly 120

Qy    457 AAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGC 516
Db    121 SerProGluValGlySerLeuArgProAla***Pro------ThrTrpArg--------- 135

Qy    517 ATCTATAATCCTACTACCCAA---------------------AGATAACCAGTTACATAT 555
Db    136 ------AsnProIleSerThrLysAsnIleLysLeuAlaGlyArgGlyGlyAlaCysLeu 153

Qy    556 TCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAG 615
Db    154 ***SerGlnLeuLeuGly---------------------------------ArgLeuArg 162

Qy    616 GCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTT 675
Db    163 GlnGluArgIleAla***ThrGlnGluAlaLysPheAlaValArgLeuCysHisArg-Th 182

Qy    676 TAAAATACATGCTTTAAAGTAGAGAAACTAGGTTG-----------------------G 711
Db    182 rProAlaTrpAlaThrArg-AlaLysLeuArgLeuLysLysLysLys*********ProG 202

Qy    712 GCATGGTGGCTCACGCCTGTATCCCA---GCACTTTGGGAGGCCGAGGCAAATGGATCAC 768
Db    202 lyValValAlaHisAlaCysAsnProSerAlaLeu-GlyValGlnGlyGlyAlaAspHis 221
```

```
Qy        769 TTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAACACCCTCTCTAATAAAA 828
              ||||||   ||||||   ||||||      ||||||   ||||||||||||||:::|||
Db        222 LeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGluThrProSerLeuLeuLys 241

Qy        829 ATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCCAGCTACTCAGGA 883
              |||    |||    ||||||:::||||||      |||:::|||||||||||||||
Db        242 Ile-***Asn***LeuGlyValValAlaHisAlaCysAsnProSerTyrSerGly 259

RESULT 11
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked re
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN:
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.:              1.02e-17        Length:         100
Score:                  232.50          Matches:        51
Percent Similarity:     75.3%           Conservative:   10
Best Local Similarity:  63.0%           Mismatches:     18
Query Match:            14.2%           Indels:         2
DB:                     2               Gaps:           1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x A46010 (1-100)

Qy        322 TTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGG 263
              |||   :::|||||||||   ||||||      ||||||:::|||||||||||||||
Db          2 PhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCys 21

Qy        262 GATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTC 203
              |||||||||||||||:::      ||| ||||||||||||:::|||   ||||||||||
Db         22 AspLeuGlySerLeuLysSerProPro-ProGlySerSerAspSerProAlaSerAlaSe 41

Qy        202 CCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGT 143
              ||||||||||||   ||||||||||||||:::||||||||||||||||||||||||||
Db         41 rArgValAlaGlyIleThrGlyMetHisHisHisThrGlnLeuIlePheValPheLeuVa 61

Qy        142 AGAGATAGGGTTTCAC---CATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGA 86
              ||||   |||   |||      :::      :::   |||   :::   |||:::   ::
Db         61 lGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAs 81

Qy         85 T 85
              :
Db         81 n 81

RESULT 12
S65657
alpha-1C-adrenergic receptor splice form 2 - human
N;Alternate names: alpha-1C-adrenoceptor isoform 2
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 09-Jul-2004
C;Accession: S65657; S65655
R;Tanaka, T.
submitted to the EMBL Data Library, July 1994
A;Reference number: S65656
A;Accession: S65657
A;Molecule type: mRNA
A;Residues: 1-499 <TAN>
A;Cross-references: UNIPROT:Q13675; UNIPARC:UPI000004E75E; EMBL:D32202; NID:g927208; PID
R;Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.;
FEBS Lett. 363, 256-260, 1995
A;Title: Cloning, functional expression and tissue distribution of human alpha(1C)-adren
A;Reference number: S65654; MUID:95255557; PMID:7737411
A;Accession: S65655
A;Molecule type: mRNA
A;Residues: 424-499 <HIR>
A;Cross-references: UNIPARC:UPI00001778AB; EMBL:D32202
C;Genetics:
A;Gene: GDB:ADRA1C; ADRA1L1
A;Cross-references: GDB:128088; OMIM:104221
A;Map position: 8p21-8p11.2
C;Superfamily: vertebrate rhodopsin
C;Keywords: alternative splicing; neurotransmitter receptor

Alignment Scores:
Pred. No.:              1.3e-15         Length:         499
Score:                  214.50          Matches:        46
Percent Similarity:     72.5%           Conservative:   4
Best Local Similarity:  66.7%           Mismatches:     14
Query Match:            13.1%           Indels:         5
DB:                     2               Gaps:           1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x S65657 (1-499)

Qy        282 AGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGC 223
              ||||||||||||||||||   |||   ||||||||||||||   ||||||||||||
Db    .   431 ArgLeuGluCysSerGlyMetIleLeuAlaHisCysAsnLeuArgLeuProGlySerArg 450

Qy        222 AATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACA---------------GGCATG 178
              :::|||   ||||||||||:::   ||||||||||||                |||
Db        451 AspSerProAlaSerAlaSerGlnAlaAlaGlyThrThrGlyAspValProProGlyArg 470

Qy        177 CATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGAC 118
                |||      |||||||||||||||||||||||||   ||||||||||||:::
Db        471 ArgHisGlnAlaGlnLeuIlePheValPheLeuValGluThrGlyPheHisHisValGly 490

Qy        117 CAGGCTGGTCTCGAACTTCTCACCTCA 91
              |||      |||:::||||||||||||
Db        491 GlnAspAspLeuAspLeuLeuThrSer 499

RESULT 13
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI0000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              6.47e-12        Length:         579
Score:                  182.00          Matches:        82
Percent Similarity:     33.0%           Conservative:   12
Best Local Similarity:  28.8%           Mismatches:     55
Query Match:            11.1%           Indels:         138
DB:                     4               Gaps:           4

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x D40201 (1-579)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||   ||||||           ||||||||||||||||||   |||:::
Db        321 AspProProAlaProAlaPhe***ValAlaGlyThrThrGlyLeuCysTyrGlnIleTrp 340
```

```
Qy        841 CTAATTTTTGTATT----TTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCT 786
              ||||||||||||      |||   |||   |||||||||:::||||||||||||   |||
Db        341 LeuIlePheValSerPhePheCysArgGlyGlyValSerThrCysCysProGlyCysSer 360

Qy        785 CAAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGC 726
              ||||||:::|||:::         |||   |||   :::      |||:::|||||
Db        361 ***ThrProGluLeuGlu-HisThrProAlaMetAlaProLysValLeuGlyLeuGlnAl 380

Qy        725 GTGAGCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGG 666
              |
Db        380 a---------------------------------------------------------- 380

Qy        665 AACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGG 606

Db        380 ---------------------------------------------------------- 380

Qy        605 AAATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACT 546
                                                    |||
Db        381 ------------------------------CysSerHisAlaTrp------------- 385

Qy        545 GGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGT 486

Db        385 ---------------------------------------------------------- 385

Qy        485 AGAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAG 426

Db        385 ---------------------------------------------------------- 385

Qy        425 ATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTT 366

Db        385 ---------------------------------------------------------- 385

Qy        365 CAGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGA 306
                                                                      |||::
Db        386 ----------------------------------------------------*********PheGl 390

Qy        305 GATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCA 246
              :      |||      ||||||||||||||||||||||      :::   |||||||| :
Db        390 nAlaGlySerTrpTyrValAlaGlnAlaGlyValGlnArgCysAsnHisGlySerLeu-S 410

Qy        245 ACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTA 186
              ::|||      |||||| |||   ||||||||||||||||||   |||   |||||||||
Db        410 erLeuAspPheProGly-GlnValIleLeuLeuProGlnProPheGlu***LeuGlyLeu 429

Qy        185 CAGGCATGCATCACTACACCCAGCTAATTTTTGTATTT----TTAGTAGAGATAGGGTTT 130
              ||||||      |||              |||||||||         ||||||:::||||||
Db        430 GlnAlaTyrAlaThrArgSerGly***PheLeuTyrLeuPhePheValGluValGlyPhe 449

Qy        129 CACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGAC 70
                 |||:::   ||||||   ||||||||||||:::   ||| ||||||   ||||||
Db        450 ArgHisValAlaGlnAlaValLeuGluLeuLeuSerTrpSer-IleHisLeuProTrpLe 469

Qy         69 TTCCAAA 63
                  |||
Db        469 uLeuLys 471
```

RESULT 14
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:

Pred. No.:	5.97e-11	Length:	579
Score:	173.50	Matches:	44
Percent Similarity:	66.2%	Conservative:	7
Best Local Similarity:	57.1%	Mismatches:	23
Query Match:	10.6%	Indels:	4
DB:	4	Gaps:	1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x D40201 (1-579)

```
Qy         89 CTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAA 148
              |||::::::|||||||||||||   ||||||:::         ||||||   |||||||||
Db        214 LeuGlnLeuArgSerSerArgThrAlaTrpAlaThrCysArgAsnProThrSerThrLys 233

Qy        149 AA----TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 204
                  |||||||||         :::|||       ||||||||||||||||||||||
Db        234 LysArgTyrLysAsn***ProAspLeuValAla***AlaCysSerProSerTyrSerLys 253

Qy        205 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 264
              |||   |||:::||||||         ||||||:::   |||||| |||||||||      |
Db        254 GlyTrpGlyArgArgIleThr-***ProGlyLysSerArgLeu-SerGluPro***LeuH 273

Qy        265 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAA 311
              ||   ||||||||||||||||||||   :::|||         |||
Db        273 isLeuCysThrProAlaTrpAlaThrTyrGlnAspProAla***Lys 288
```

RESULT 15
I54375
gene NF2 protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Oct-2004
C;Accession: I54375
R;Arakawa, H.; Hayashi, N.; Nagase, H.; Ogawa, M.; Nakamura, Y.
Hum. Mol. Genet. 3, 565-568, 1994
A;Title: Alternative splicing of the NF2 gene and its mutation analysis of breast and co
A;Reference number: I54375; MUID:94348501; PMID:8069299
A;Accession: I54375
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-46 <RES>
A;Cross-references: UNIPROT:Q16230; UNIPARC:UPI000007034D; GB:S73853; NID:g688372; PIDN
C;Genetics:
A;Gene: GDB:NF2
A;Cross-references: GDB:120232; OMIM:101000
A;Map position: 22q12.2-22q12.2

Alignment Scores:

Pred. No.:	7.07e-11	Length:	46
Score:	172.00	Matches:	32
Percent Similarity:	79.1%	Conservative:	2
Best Local Similarity:	74.4%	Mismatches:	9
Query Match:	10.5%	Indels:	0
DB:	2	Gaps:	0

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x I54375 (1-46)

```
Qy        310 TTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCA 251
              |||      ||||||      |||      ||||||||||||||   |||||||||
Db          4 PheAsnCysGluSerCysSerValThrLeuAlaGlyValGlnTrpArgAspLeuGlyLeu 23

Qy        250 CTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGG 191
              ||||||||||||||||||:::||||||:::|||||||||||||||   ||||||||||||
Db         24 LeuGlnProLeuProProLysPheLysArgPheSerCysLeuSerPheProSerSerTrp 43

Qy        190 GACTACAGG 182
              |||||||||
```

Db 44 AspTyrArg 46

Search completed: February 8, 2006, 11:19:08
Job time : 39.7434 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31 ; Search time 41.5475 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-T9182_COPY_8700_9600
Perfect score: 1640
Sequence: 1 tgttcatattaattaaaaat.....gagtctgaggcaggagtatc 901

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_694_@runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	493.5	30.1	593	1	ALU7_HUMAN	P39194 homo sapien
c	2	471	28.7	593	1	ALU6_HUMAN	P39193 homo sapien
c	3	445	27.1	591	1	ALU8_HUMAN	P39195 homo sapien
c	4	437.5	26.7	587	1	ALU2_HUMAN	P39189 homo sapien
c	5	436	26.6	591	1	ALU1_HUMAN	P39188 homo sapien
c	6	427.5	26.1	587	1	ALU3_HUMAN	P39190 homo sapien
	7	427.5	26.1	593	1	ALU7_HUMAN	P39194 homo sapien
c	8	422	25.7	585	1	ALU5_HUMAN	P39192 homo sapien
	9	411	25.1	591	1	ALU1_HUMAN	P39188 homo sapien
	10	400	24.4	585	1	ALU5_HUMAN	P39192 homo sapien
	11	400	24.4	593	1	ALU6_HUMAN	P39193 homo sapien
	12	394	24.0	591	1	ALU8_HUMAN	P39195 homo sapien
c	13	391.5	23.9	179	2	Q6ZSI5_HUMAN	Q6zsi5 homo sapien
c	14	389.5	23.8	375	2	O60448_HUMAN	O60448 homo sapien
	15	371.5	22.7	587	1	ALU2_HUMAN	P39189 homo sapien
c	16	368	22.4	164	2	Q6ZTS0_HUMAN	Q6zts0 homo sapien
c	17	366.5	22.3	156	2	Q8NBH4_HUMAN	Q8nbh4 homo sapien
c	18	361	22.0	146	2	Q6ZRX6_HUMAN	Q6zrx6 homo sapien
c	19	360.5	22.0	128	2	Q6ZPB2_HUMAN	Q6zpb2 homo sapien
c	20	359	21.9	239	2	Q6ZWA9_HUMAN	Q6zwa9 homo sapien
c	21	350.5	21.4	129	2	Q6ZUW8_HUMAN	Q6zuw8 homo sapien
c	22	348	21.2	170	2	Q8NAI3_HUMAN	Q8nai3 homo sapien
c	23	344.5	21.0	150	2	Q6ZPA0_HUMAN	Q6zpa0 homo sapien
c	24	344	21.0	123	2	Q6ZMQ3_HUMAN	Q6zmq3 homo sapien
c	25	342	20.9	129	2	Q8N9K0_HUMAN	Q8n9k0 homo sapien
	26	339.5	20.7	587	1	ALU3_HUMAN	P39190 homo sapien
c	27	336	20.5	124	2	Q6ZSN0_HUMAN	Q6zsn0 homo sapien
c	28	334	20.4	164	2	Q6ZUK0_HUMAN	Q6zuk0 homo sapien
c	29	333	20.3	137	2	Q6ZS97_HUMAN	Q6zs97 homo sapien
c	30	331.5	20.2	127	2	Q6ZVW2_HUMAN	Q6zvw2 homo sapien
c	31	329	20.1	167	2	Q6ZSN3_HUMAN	Q6zsn3 homo sapien
c	32	328.5	20.0	120	2	Q6ZTA3_HUMAN	Q6zta3 homo sapien
c	33	328	20.0	140	2	Q6ZP99_HUMAN	Q6zp99 homo sapien
c	34	327.5	20.0	115	2	Q9N083_MACFA	Q9n083 macaca fasc
c	35	327	19.9	131	2	Q6ZTD3_HUMAN	Q6ztd3 homo sapien
c	36	325	19.8	132	2	Q6ZS28_HUMAN	Q6zs28 homo sapien
c	37	324	19.8	157	2	Q8N845_HUMAN	Q8n845 homo sapien
c	38	324	19.8	202	2	Q6ZUA4_HUMAN	Q6zua4 homo sapien
c	39	324	19.8	204	2	Q8N8P8_HUMAN	Q8n8p8 homo sapien
c	40	324	19.8	299	2	Q6ZRN6_HUMAN	Q6zrn6 homo sapien
c	41	323.5	19.7	166	2	Q6ZUC0_HUMAN	Q6zuc0 homo sapien
c	42	323	19.7	145	2	Q4R5J6_MACFA	Q4r5j6 macaca fasc
	43	323	19.7	603	1	ALU4_HUMAN	P39191 homo sapien
c	44	322	19.6	132	2	Q6ZNU7_HUMAN	Q6znu7 homo sapien
c	45	321.5	19.6	139	2	Q6ZUL7_HUMAN	Q6zul7 homo sapien

ALIGNMENTS

RESULT 1
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:              6.25e-46       Length:         593
Score:                  493.50         Matches:        114
Percent Similarity:     43.4%          Conservative:   14
Best Local Similarity:  38.6%          Mismatches:     40
Query Match:            30.1%          Indels:         128
DB:                     1              Gaps:           2

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU7_HUMAN (1-593)

Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
              :::|||||||||   |||   ||||||||||||||||||      ||||||   ||||||

Db        331 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProProArgProAla 350

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
              ||||||||||||||   |||:::|||||||||   |||   ||||||||||||:::|||

Db        351 AsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerArgThr 370

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGTGAG 721
              |||||||||:::   |||      ||||||||||||||||||||   ||||||||||||

Db        371 ProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspTyrArgArgGlu 390

Qy        720 CCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACAT 661
              ||||||   |||            :::|||

Db        391 ProProArgProAla*********PhePhe***-------------------------- 401

Qy        660 CCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATT 601

Db        401 ---------------------------------------------------------- 401

Qy        600 TTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTA 541

Db        401 ---------------------------------------------------------- 401

Qy        540 TCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGT 481

Db        401 ---------------------------------------------------------- 401

Qy        480 TTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTAC 421

Db        401 ---------------------------------------------------------- 401

Qy        420 AGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGA 361
                                                                      |||

Db        402 ----------------------------------------------------------AspGly 403

Qy        360 ATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAG 301
              :::   |||:::|||

Db        404 ValSerLeuLeuLeu---------------------------------------------- 408

Qy        300 AGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTC 241
                          ||||||||||||||||||||   ||||||||||||||||||||

Db        409 --------------ProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeu 423

Qy        240 TGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGC 181
                    ||||||||||||||:::|||   ||||||||||||||||||||||   ||||||

Db        424 ArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGly 443

Qy        180 ATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATT 121
                    |||:::   :::|||||||||||||||||||||||   ||||||||||||:::

Db        444 AlaArgHisHisAlaArgLeuIlePheValPheLeuValGluThrGlyPheHisHisVal 463

Qy        120 GACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGT 61
                    ||||||||||||||||||||||||||   |||||||||           ||||||

Db        464 GlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSer 483

Qy         60 ACTGTGGTTACAGGCTTGAGTCACTGTGCCCAACTGAGTTATTTT 16
                    :::||||||:::||||||   |||:::          |||

Db        484 AlaGlyIleThrGlyValSerHisArgAlaArg*********Phe 498

RESULT 2
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
```

```
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              2.13e-43        Length:         593
Score:                  471.00          Matches:        130
Percent Similarity:     53.0%           Conservative:   27
Best Local Similarity:  43.9%           Mismatches:     100
Query Match:            28.7%           Indels:         40
DB:                     1               Gaps:           8

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU6_HUMAN (1-593)

Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
              :::|||||||||   |||   ||||||||||||||||||      ||||||   ||||||
Db        331 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArgProAla 350

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
              ||||||||||||||   |||:::|||||||||   |||      |||||||||:::|||
Db        351 AsnPheCysIlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSerArgThr 370

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGTGAG 721
              |||||||||:::   |||      ||||||||||||||||||||||   ||||||||||||
Db        371 ProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGlu 390

Qy        720 CCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACAT 661
              ||||||   |||           :::|||          :::      ::::::
Db        391 ProProArgProAla*********PhePhe***AspGlyValSerLeuLeuLeuProArg 410

Qy        660 CCTTAAATGGAAT---ACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAA 604
              |||:::   |||                  |||   :::|||
Db        411 -LeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArg---------------- 424

Qy        603 ATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGG 544
                      :::      :::|||      |||   :::            |||   ||
Db        425 ----------LeuProGlySerSerAspSerProAlaSerAlaSerArg---ValAlaGl 440

Qy        543 TTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAG 484
              |         :::   |||:::                        |||:::      ||
Db        440 y---------IleThrGlyMet----------------------ArgHisHisAlaAr 449

Qy        483 AGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGAT 424
              |:::                  :::|||   |||:::   :::
Db        449 gLeu--------------------IlePheValPheLeuValGlu-------------- 457

Qy        423 TACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCT---------- 375
              |||      :::|||:::                  |||      |||
Db        458 -ThrGlyPheLeuHisValGlyGlnAlaGlyLeuGluLeuProThrSerGlyAspProPr 477

Qy        374 -GCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTT 316
              |||         :::|||
Db        477 oAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg******** 497

Qy        315 ATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCG 256
              ||||||                  |||   |||||||||||||:::      |||||
Db        497 *PhePheGluThrGluPheArgSerCysCysProGlyTrpSerAlaMetAlaArgSerAr 517

Qy        255 GCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTA 196
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        517 gLeuThrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProGlu** 537

Qy        195 GCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATA 136
              ||||||||||||||||||   |||||||||      ||||||||||||      |||
Db        537 *LeuGlyLeuGlnAlaCysAlaThrThrProGly***PheLeuTyrPhe******ArgAr 557

Qy        135 GGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACC 76
              ||||||::::::|||   |||||||||||||||      ||||||||||||   |||||
Db        557 gGlyPheSerMetLeuValArgLeuValSerAsnSerArgProGlnValIleArgProPr 577

Qy         75 TTGGACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCACTGTGCCC 30
              |         |||||||||   |||||||||      |||   |||
Db        577 oArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaPro 592

RESULT 3
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
```

```
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:              1.79e-40       Length:         591
Score:                  445.00         Matches:        128
Percent Similarity:     46.1%          Conservative:   25
Best Local Similarity:  38.6%          Mismatches:     66
Query Match:            27.1%          Indels:         113
DB:                     1              Gaps:           10

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU8_HUMAN (1-591)
```

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||   |||||||||   |||||||||:::
Db        329 AspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisHisAlaArg 348

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
              |||||||||||||||||:::|||      ||||||||||||   ||||||||||||:::
Db        349 LeuIlePheValPheLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGlu 368

Qy        781 CTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGA 722
              |||   ||||||   |||                  :::      |||:::||||||
Db        369 LeuLeuThrSerGlyAsp---------LeuProProLysValLeuGlyLeuGlnAla*** 385

Qy        721 GCCACCATGCCCAACCTA----------------------------GTTTCTCTACTT 692
              ||||||   |||   |||                              |||||||||
Db        386 AlaThrAlaProGlyLeuArgPro*********PhePhe***AspGlyValSerLeuCys 405

Qy        691 ---------TAAAGCATGTATTTTAAAAATTATGGAACATCCTTAAATGGA--------- 650
                      |||         ::::::      |||:::   :::
Db        406 ArgProGlyTrpSerAlaValAlaArgSerArgLeuThrAlaThrSerAlaSerArgVal 425

Qy        649 ---------ATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTTT 599
                      :::|||   :::      |||   ||||||:::   |||
Db        426 GlnAlaIleLeuLeuProGlnProProGlu***LeuGlyLeuGlnAlaArgAlaThrThr 445

Qy        598 CCC------------------------------------------------------- 596
              |||
Db        446 ProGly***PheLeuTyrPhe******ArgArgGlyPheThrMetLeuAlaArgLeuVal 465

Qy        595 ------AAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTT 542
                        ::::::::::::   ||||||   |||
Db        466 SerAsnSer***ProGlnValIlePheLeuProLysCysTrp----------------- 479

Qy        541 ATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAG 482
                          |||||||||      |||
Db        480 --------------AspTyrArgArgGluPro-------------------------- 485

Qy        481 TTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTA 422
                                                  |||:::
Db        486 ----------------------------------------ProArg------------- 487

Qy        421 CAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGG 362

Db        487 ------------------------------------------------------------ 487

Qy        361 AATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATA 302
                                  |||   |||                  |||||||||
Db        488 ------------ProAlaSerAlaArg***------------******PhePheGluThr 499

Qy        301 GAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCT 242
              |||||||||   ||||||||||||||||||||||||   ||||||||||||||||||||
Db        500 GluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnPro 519

Qy        241 CTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGG 182
                ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||
Db        520 ProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArg 539

Qy        181 CATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATAT 122
                |||         ||||||||||||||||||||||||||||||   |||||||||
Db        540 ArgAlaProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerProCys 559

Qy        121 TGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAG 62
              |||||||||||||||||      |||:::   |||::: :::         |||||
Db        560 TrpProGlyTrpSerArgThrProAspLeuArg***SerSer-Ser---------GlnSe 576

Qy         61 TACTGTGGTTACAGGCTTGAGTCACTGTGCCCAA 28
              |      :::||||||:::||||||   |||:::
Db        576 rAlaGlyIleThrGlyValSerHisArgAlaArg 587
```

```
RESULT 4
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
```

```
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:              1.25e-39      Length:          587
Score:                  437.50        Matches:         133
Percent Similarity:     50.0%         Conservative:    16
Best Local Similarity:  44.6%         Mismatches:      106
Query Match:            26.7%         Indels:          44
DB:                     1             Gaps:            6

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU2_HUMAN (1-587)

Qy        898 ACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCTA 839
              :::|||||||||   |||   |||||||||||||||||||||   |||||||||:::|||
Db        328 SerProAlaSerAlaSerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArgLeu 347

Qy        838 ATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACTC 779
              ||||||||||||||:::|||      ||||||   |||   |||   ||||||   |||
Db        348 IlePheValPheLeuValGluThrGlyPheHisArgValSerGlnAspGlyLeuAspLeu 367

Qy        778 CGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGTGAGC 720
              ||||||    |||      ||||||||||||||||||||   ||||||||||||
Db        368 LeuThrSer***--SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluP 387

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTT-----------AAAGCATGTATTTTAAAAA 672
              |||||   |||          :::|||                |||
Db        387 roProArgProAla*********PhePhe***AspGlyValSerLeuCys---------- 403

Qy        671 TTATGGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGC 612
                        |||    |||:::                      |||
Db        404 --------ArgPro-GlyTrpSerAlaVal--------------AlaArgSerArgLeu 415

Qy        611 ACATGGAAATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATAT 552
              |||         :::            |||   :::   |||:::      |||
Db        416 ThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProProGlu------ 433

Qy        551 GTAACTGGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATAT 492
                              :::||||||      |||:::|||
Db        434 ------------------***LeuGlyLeuGlnAlaProAlaThr--------------- 442

Qy        491 TTATGTAGAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAAC 432
                                  |||   |||   |||   :::
Db        443 ----------------------ThrProGly***PheLeuTyrPhe***---***Arg 453

Qy        431 AATAAGATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCT 372
                       ||||||      :::     |||:::                  |||
Db        454 ArgGlyPheThrValLeuAlaArgMetValSerIleSer***ProArgAspProProAla 473

Qy        371 TACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTT 312
                  :::|||               :::                          ||
Db        474 SerAlaSerGlnSer-AlaGlyIleThrGlyValSerHisArgAlaArg*********Ph 493

Qy        311 TTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTC 252
              |||||||   |||||||||   |||||||||||||||||||||||||   ||||||||||
Db        493 ePheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySe 513

Qy        251 ACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTG 192
              |||||||     ||||||  |||     |||||||||||||||||||||||||||||||
Db        513 rLeuGlnAlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTr 533

Qy        191 GGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGT 132
              ||||||||||                |||||||||||||||||||||||||||||   ||
Db        533 pAspTyrArgArgProProProArgProAlaAsnPheCysIlePheSerArgAspGlyVa 553

Qy        131 TTCACCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGG 72
              |||||||      |||||||||||||::      |||  ||||||   ||||||
Db        553 lSerProCys***ProGlyTrpSerArgSerProAspLeu--ValIleArgProProArg 572
```

```
Qy         71 ACTTCCAAAGTACTGTGGTTACAGGCTTGAGTCACTGTGCCC 30
                ||||||||| |||||||||     |||   |||
Db        573 ProProLysValLeuGlyLeuGlnAla***AlaThrAlaPro 586

RESULT 5
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:              1.84e-39      Length:         591
Score:                  436.00        Matches:        131
Percent Similarity:     47.9%         Conservative:   18
Best Local Similarity:  42.1%         Mismatches:     87
Query Match:            26.6%         Indels:         75
DB:                     1             Gaps:           7

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU1_HUMAN (1-591)

Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||   |||||||||   |||   |||||||||||||||||||||   |||||||||:::
Db        329 AspProProAlaSerAlaSerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArg 348

Qy        841 CTAATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAA 782
              |||||||||||||||   :::|||      ||||||:::|||   |||||||||||:::
Db        349 LeuIlePheValPhePheValGluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGlu 368

Qy        781 CTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGA 722
              |||       ||||||||||||   |||||||||||||||||||||||||||
Db        369 LeuLeuGlySerSerAspProProAlaSerAlaSerGlnSerAlaGlyIle--------- 385

Qy        721 GCCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACA 662

Db        385 ------------------------------------------------------------ 385

Qy        661 TCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAAT 602
                  |||:::   ||||||
Db        386 ------ThrGlyValSerHisArg-----------------AlaArg***-******Ph 397

Qy        601 TTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTT 542
               |||   ::::::   |||   |||   |||                       |||
Db        397 ePhe***AspArgValSerLeuCysArgPro--------------------GlyTrpSe 410

Qy        541 ATCTTTGGGT---AGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTA 485
               ::::::      ||||||:::      ||| ||||||  |||
Db        410 rAlaValAlaArgSerArgLeuThrAlaAla-SerThr-SerArgAlaGln--------- 426

Qy        484 GAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGA 425
                                  |||   ||||||      |||        :::
Db        427 ----------------------AlaIleLeuLeuProGlnProProGlu***LeuGly 438

Qy        424 TTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTC 365
              ||||||   :::                  :::   ::::::           |||
Db        439 LeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu***ArgArgGlyPhe 458

Qy        364 AGGAATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATT------------ 317
                 |||       ||||||      |||:::       |||
Db        459 ThrMetLeuProArgLeuValSer---AsnSerTrpAlaGlnValIleLeuProProArg 477

Qy        316 ------------------------------------------------TATTTTTTT 308
                                                             |||||||
Db        478 ProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly*********PhePhe 497

Qy        307 GAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTG 248
              |||       ||||||   |||||||||||||||||||||||   |||   |||||||||
Db        498 GluThrGlySerArgSerValAlaGlnAlaGlyValGlnTrpArgAspHisGlySerLeu 517
```

```
Qy        247 CAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGAC 188
              ||||||  ||||||     |||:::  ||||||||||||||||||||||||||||||
Db        518 GlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuProSerSerTrpAsp 537

Qy        187 TACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCA 128
              ||||||  |||         ||||||||||||||||||||  ||||||  ||||||
Db        538 TyrArgArgAlaProProArgProAlaAsnPheCysIlePheCysArgAspGlyValSer 557

Qy        127 CCATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTT 68
              ||||||||||||||||||      ||||||  |||:::  |||||||||
Db        558 LeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSerArgLeuGlyLeu 577

Qy         67 CCAAAGTACTGTGGTTACAGGCTTGAG 41
              ||||||      ||||||  |||
Db        578 ProLysCysTrpAspTyrArgArgGlu 586
```

```
RESULT 6
ALU3_HUMAN
ID   ALU3_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39190;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB1 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14569; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63613 MW;  85C4155726DEF235 CRC64;
```

```
Alignment Scores:
Pred. No.:              1.67e-38        Length:         587
Score:                  427.50          Matches:        131
Percent Similarity:     48.4%           Conservative:   18
Best Local Similarity:  42.5%           Mismatches:     96
Query Match:            26.1%           Indels:         64
DB:                     1               Gaps:           7

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU3_HUMAN (1-587)

Qy        898 ACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCTA 839
              :::|||||||||   |||   ||||||||||||||||||||   |||:::|||:::|||
Db        328 SerProAlaSerAlaSerGlnValAlaGlyThrThrGlyAlaArgHisTyrArgAlaLeu 347

Qy        838 ATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACTC 779
              |||||||||||||||:::|||      ||||||         :::   ||||||  |||
Db        348 IlePheValPheLeuValGluThrGlyPheHisArgPheSerArgAspGlyLeuAspLeu 367

Qy        778 CGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGA-TACAGGCGTGAGC 720
              ||||||     |||      ||||||||||||||||||||||   ||||||||||||
Db        368 LeuThrSer***--SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluP 387

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTT-----------AAAGCATGTATTTTAAAAA 672
              |||||   |||               :::|||                |||
Db        387 roProArgProAla*********PhePhe***AspGlyValSerLeuCys---------- 403

Qy        671 TTATGGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGC 612
                      |||    |||:::
Db        404 --------ArgPro-GlyTrpSerAlaVal---------------------------- 410

Qy        611 ACATGGAAATTTTCCCAAATAAAAAAAGCTAGTGTATGC-------------------- 573
                      |||::::::      :::|||
Db        411 ------AlaGlySerArgLeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeu 428

Qy        572 ---------CCCAAAAATCTGGAGGAATATGTAACTGGTTATCTTTGGGTAGTAGGATTA 522
                      ||||||  |||                                  ||||||
Db        429 ProGlnProProLys***Leu--------------------------------GlyLeu 437

Qy        521 TAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACTTTT 462
                 |||:::|||                                        |||
Db        438 GlnAlaProAlaThr-----------------------------------------ThrPro 444

Qy        461 GGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAAC 402
              |||  |||  :::                                 ||||||      :::
Db        445 Gly***PheLeuTyrPhe***---***ArgArgGlyPheThrValLeuAlaGlyMetVal 463
```

```
Qy        401 TTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTC 342
                 |||:::                  |||         :::|||
Db        464 SerIleSer***ProArgAspProProAlaSerAlaSerGlnSer-AlaGlyIleThrGl 483

Qy        341 TCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCA 282
                :::                      |||||||||   |||||||||   ||||||||
Db        483 yValSerHisArgAlaArg********PhePheGluThrGluSerArgSerValAlaGl 503

Qy        281 GGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCA 222
                |||||||||||||||   ||||||||||||||||||      ||||||   |||
Db        503 nAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyPheThrPr 523

Qy        221 ATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGC 162
                ||||||||||||||||||||||||||||||||||||||||          |||   |||||
Db        523 oPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProLeuArgProAl 543

Qy        161 TAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAAC 102
                ||||||||||||||||||||||   |||||||||:::   ||||||||||||||::
Db        543 aAsnPheCysIlePheSerArgAspGlyValSerProPhe***ProGlyTrpSerArgSe 563

Qy        101 TTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCTTGA 42
                :      |||  ||||||   ||||||         |||||||||   |||||||||
Db        563 rProAspLeu--ValIleArgProProArgProProLysValLeuGlyLeuGlnAla*** 582

Qy         41 GTCACTGTGCCC 30
                 |||   |||
Db        583 AlaThrAlaPro 586
```

RESULT 7
ALU7_HUMAN
ID   ALU7_HUMAN   STANDARD;   PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC   into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC   consensus sequences have been constituted that contain all six
CC   frames conceptual translations of each of these classes of Alu
CC   repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC   codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC   pollution of protein sequence databases with Alu-derived amino
CC   acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC   primate genomes with an average spacing of 4 kb. Some of them are
CC   actively transcribed by pol III. Normal transcripts may contain
CC   Alu-derived sequences in 5' or 3' untranslated regions. However,
CC   cDNA libraries also contain partial and/or rearranged cDNAs
CC   ligated with Alu-derived sequence in any orientation. Although Alu
CC   elements (especially situated on the complementary strand) have a
CC   great potential to create additional/alternative exons,
CC   consideration should be given to the possibility that the presence
CC   of an Alu in an open reading frame may have resulted from a
CC   cloning artifact or may be due to misinterpretation of sequencing
CC   data. This point has been overlooked on several occasions, with
CC   the consequence of erroneous Alu-derived amino acid sequences
CC   being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC   with an Alu-translated entry must be taken as a warning that a
CC   part of Alu repeat may have been artifactually included in the
CC   coding nucleotide sequence.
CC   ---
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION   1   97   Frame-1.
FT   REGION   101   196   Frame-2.
FT   REGION   200   295   Frame-3.
FT   REGION   299   395   Frame-4.
FT   REGION   399   494   Frame-5.
FT   REGION   498   593   Frame-6.
SQ   SEQUENCE   593 AA; 64455 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:

Pred. No.:	1.67e-38	Length:	593
Score:	427.50	Matches:	112
Percent Similarity:	43.3%	Conservative:	13
Best Local Similarity:	38.8%	Mismatches:	37
Query Match:	26.1%	Indels:	128
DB:	1	Gaps:	3

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU7_HUMAN (1-593)

```
Qy         32 GCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTT 91
                ||:::   |||   ||||||         ||||||:::::::::   |||||||||
Db          3 AlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySerPro 22

Qy         92 GAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAAT 151
                |||||||||||||||||||||||||     |||   ||||||:::||||||||||||
Db         23 GluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThrLysAsn 42

Qy        152 ACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG 211
                ||||||||||||                 ||||||:::||||||||||||||||||||
Db         43 ThrLysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGlu 62

Qy        212 GCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTG 271
                |||   |||   ||||||||||||||:::|||||||||   |||||||||   ||||||
Db         63 AlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeu 82

Qy        272 CACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 331
                ||||||||||||||
Db         83 HisSerSerLeuGly---------------------------------------------- 87

Qy        332 GGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTAT 391
                      ||||||   ::::::   |||      |||:::
```

```
Db         88 ------------AsnLys---SerGluThrProSerGlnLysLys*********------ 100

Qy        392 ATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAG 451

Db        100 ------------------------------------------------------------ 100

Qy        452 TAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGG 511

Db        100 ------------------------------------------------------------ 100

Qy        512 AAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGG 571

Db        100 ------------------------------------------------------------ 100

Qy        572 GGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTC 631

Db        100 ------------------------------------------------------------ 100

Qy        632 TAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTA 691

Db        100 ------------------------------------------------------------ 100

Qy        692 AAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGG 750
                          |||   ||||||||||||||||   ||||||||||||||||||
Db        101 --------------AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArg 115

Qy        751 CCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAA 810
              ||||||::::::|||||||||||||||||||||   ||||||||||||||||||   |||
Db        116 ProArgArgValAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGlu 135

Qy        811 ACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870
              ||||||||||||:::|||||||||||||||||||   ||||||      |||   |||
Db        136 ThrProSerLeuLeuLysIleGlnLysLeuAlaGlyArgGlyGlyGlyArgLeu***Ser 155

Qy        871 CAGCTACTCAGGAGTCTGAGGCAGGAG 897
              |||||||||      ||||||||||||
Db        156 GlnLeuLeuGlyArgLeuArgGlnGlu 164
```

```
RESULT 8
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.:              6.92e-38      Length:          585
Score:                  422.00        Matches:         113
Percent Similarity:     41.7%         Conservative:    8
Best Local Similarity:  39.0%         Mismatches:      36
Query Match:            25.7%         Indels:          134
DB:                     1             Gaps:            1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU5_HUMAN (1-585)
```

```
Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
              :::|||||||||   |||   ||||||||||||||||||      ||||||   ||||||
Db        424 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgAlaProProArgProAla 443

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
              ||||||||||||||   |||:::||||||||   |||   ||||||||||||:::::
Db        444 AsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerArgSer 463

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATA-CAGGCGTGAG 721
                  ||||||  ||||||      ||||||||||||||||||||||||   ||||||   |
Db        464 LeuAspLeu--ValIleArgProProArgProProLysValLeuGlyLeuGlnAla***A 483

Qy        720 CCACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACAT 661
              |||||   |||
```

```
Db        483 laThrAlaProGly*********----------------------------------- 490

Qy        660 CCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATT 601

Db        490 ------------------------------------------------------------ 490

Qy        600 TTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTA 541

Db        490 ------------------------------------------------------------ 490

Qy        540 TCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGT 481

Db        490 ------------------------------------------------------------ 490

Qy        480 TTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTAC 421

Db        490 ------------------------------------------------------------ 490

Qy        420 AGTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGA 361

Db        490 ------------------------------------------------------------ 490

Qy        360 ATCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAG 301
                                                  |||||||||   |
Db        491 ----------------------------------------------PhePheGluThrG 495

Qy        300 AGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTC 241
              ||||||||   |||||| ||||||||||||||||||   ||||||||||||||||||||
Db        495 luSerArgSerValAla-ArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeu 514

Qy        240 TGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGC 181
                ||||||||||||||:::|||   ||||||||||||||||||||||||||||||||||
Db        515 ArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyThrThrGly 534

Qy        180 ATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATT 121
                  |||:::   ||||||||||||||||||||||||||   ||||||||||||:::
Db        535 AlaArgHisHisAlaGlnLeuIlePheValPheLeuValGluThrGlyPheHisHisVal 554

Qy        120 GACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGT 61
                |||   ||||||:::|||||||||||||     |||      ||||||||||||||
Db        555 GlyGlnAspGlyLeuAspLeuLeuThrSer***--SerAlaArgLeuGlyLeuProLysC 574

Qy         60 ACTGTGGTTACAGGCTTGAG 41
                      ||||||   |||
Db        574 ysTrpAspTyrArgArgGlu 580
```

```
RESULT 9
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:              1.2e-36        Length:          591
Score:                  411.00         Matches:         110
Percent Similarity:     40.5%          Conservative:    7
Best Local Similarity:  38.1%          Mismatches:      44
Query Match:            25.1%          Indels:          129
DB:                     1              Gaps:            1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU1_HUMAN (1-591)

Qy         32 GCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTT 91
              |||:::   |||   ||||||         ||||||:::::::::   ||||||||||||
Db          3 AlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySerLeu 22

Qy         92 GAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAAT 151
              |||   |||||||||||||||||||:::         ||||||:::||||||||||||
```

```
Db         23 GluProArgSerSerArgProAlaTrpAlaThr******AsnProValSerThrLysAsn 42

Qy        152 ACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG 211
              ||||||||||||                ||||||||||||||||||||||||||||||
Db         43 ThrLysIleSerArgAlaTrpTrpArgAlaProValValProAlaThrArgGluAlaGlu 62

Qy        212 GCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTG 271
              |||         ||||||||||||||   |||||||||   |||   |||   ||||||
Db         63 AlaGlyGlySerLeuGluProGlyArgSerArgLeuGln***AlaValIleAlaProLeu 82

Qy        272 CACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 331
              ||||||||||||||||||   ||||||      |||||||||
Db         83 HisSerSerLeuGlyAspArgAlaArgProCysLeuLysLys*********--------- 99

Qy        332 GGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTAT 391

Db         99 ------------------------------------------------------------ 99

Qy        392 ATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAG 451

Db         99 ------------------------------------------------------------ 99

Qy        452 TAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGG 511

Db         99 ------------------------------------------------------------ 99

Qy        512 AAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGG 571

Db         99 ------------------------------------------------------------ 99

Qy        572 GGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTC 631

Db         99 ------------------------------------------------------------ 99

Qy        632 TAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTA 691

Db         99 ------------------------------------------------------------ 99

Qy        692 AAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGG 750
                          |||   ||||||||||||||   ||||||||||||||||||
Db        100 --------------AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArg 114

Qy        751 CCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAA 810
              |||||:::   |||||||||      ||||||   ||||||||||||||||||   |||
Db        115 ProArgArgGluAspHisLeuSerProGlyValArgAspGlnProGlyGlnHisSerGlu 134

Qy        811 ACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870
              ||||||||||||   ||||||||||||||||||||   ||||||      |||   |||
Db        135 ThrProSerLeuGlnLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***Ser 154

Qy        871 CAGCTACTCAGGAGTCTGAGGCAGGAG 897
              |||||||||      ||||||||||||
Db        155 GlnLeuLeuGlyArgLeuArgGlnGlu 163
```

```
RESULT 10
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;
```

```
Alignment Scores:
Pred. No.:              2.07e-35      Length:          585
Score:                  400.00        Matches:         109
Percent Similarity:     42.1%         Conservative:    14
Best Local Similarity:  37.3%         Mismatches:      36
Query Match:            24.4%         Indels:          134
DB:                     1             Gaps:            1
```

```
US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU5_HUMAN (1-585)

Qy         32 GCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTT 91
              |||:::   |||   ||||||         ||||||::::::::   |||||||||
Db          3 AlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySerArg 22

Qy         92 GAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAAT 151
                |||::::::   ||||||   |||   ||||||||||||||   ||||||   ||||
Db         23 --GlyGlnGluIleGluThrIleLeuAlaAsnMetValLysProArgLeuTyr***LysT 42

Qy        152 ACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG 211
              ||||||||   ||||||||||||      |||||||||||||||||||||||||||   |
Db         42 yrLysAsn***LeuGlyValValAlaArgAlaCysSerProSerTyrSerGlyGly***G 62

Qy        212 GCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTG 271
              ||:::|||||||||   |||   |||||||||||||||||||||||||      ||||
Db         62 lyArgArgIleAla***ThrArgGluAlaGluValAlaValSerArgAspArgAlaThrA 82

Qy        272 CACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 331
              ||||||||||| ||||||   |||||||||   |||||||||
Db         82 laLeuGlnPro-GlyAspArgAlaArgLeuArgLeuLysLys********--------- 98

Qy        332 GGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTAT 391

Db         98 ------------------------------------------------------------ 98

Qy        392 ATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAG 451

Db         98 ------------------------------------------------------------ 98

Qy        452 TAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGG 511

Db         98 ------------------------------------------------------------ 98

Qy        512 AAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGG 571

Db         98 ------------------------------------------------------------ 98

Qy        572 GGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTC 631

Db         98 ------------------------------------------------------------ 98

Qy        632 TAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTA 691

Db         98 ------------------------------------------------------------ 98

Qy        692 AAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGG 750
                          |||   ||||||||||||||   ||||||||||||||||||
Db         99 --------------AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArg 113

Qy        751 CCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAA 810
              ||||||:::   ||||||  ||||||:::      ||||||:::|||   |||      |
Db        114 ProArgArgAlaAspHis--GluValLysArgSerArgProSerTrpProThrTrp***A 133

Qy        811 ACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870
              ||   :::|||   |||||||||||||||||||||||||||:::|||||||||||||
Db        133 snProValSerThrLysAsnThrLysIleSerTrpAlaTrpTrpArgAlaProValValP 153

Qy        871 CAGCTACTCAGGAGTCTGAGGCAGGA 896
              ||||||||:::|||:::|||||||||
Db        153 roAlaThrArgGluAlaGluAlaGly 161

RESULT 11
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
```

```
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              2.07e-35      Length:          593
Score:                  400.00        Matches:         109
Percent Similarity:     42.2%         Conservative:    13
Best Local Similarity:  37.7%         Mismatches:      40
Query Match:            24.4%         Indels:          128
DB:                     1             Gaps:            3

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU6_HUMAN (1-593)

Qy         32 GCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTT 91
              |||:::   |||   ||||||         ||||||:::::::::   |||||||||
Db          3 AlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySerPro 22

Qy         92 GAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAAT 151
              ||||||   ||||||||||||||||         |||   ||||||:::|||||||||||||
Db         23 GluValGlySerSerArgProAla***ProThrTrpArgAsnProValSerThrLysAsn 42

Qy        152 ACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG 211
              ||||||||||||                ||||||||:::||||||||||||||||||||||
Db         43 ThrLysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGlu 62

Qy        212 GCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTG 271
              |||   |||   ||||||||||||||:::|||||||:::   |||||||||   ||||||
Db         63 AlaGlyGluSerLeuGluProGlyArgArgArgLeuArg***AlaGluIleAlaProLeu 82

Qy        272 CACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 331
              ||||||||||||||
Db         83 HisSerSerLeuGly---------------------------------------------- 87

Qy        332 GGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTAT 391
                          ||||||   ::::::   |||      |||:::
Db         88 ------------AsnLys---SerGluThrProSerGlnLysLys*********------ 100

Qy        392 ATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAG 451

Db        100 ---------------------------------------------------------- 100

Qy        452 TAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGG 511

Db        100 ---------------------------------------------------------- 100

Qy        512 AAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGG 571

Db        100 ---------------------------------------------------------- 100

Qy        572 GGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTC 631

Db        100 ---------------------------------------------------------- 100

Qy        632 TAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTA 691

Db        100 ---------------------------------------------------------- 100

Qy        692 AAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTA-TCCCAGCACTTTGGGAGG 750
                          |||   ||||||||||||||   ||||||||||||||||||
Db        101 --------------AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArg 115

Qy        751 CCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAA 810
              ||||||:::   ||||||||||||||||||||   |||||||||   ||||||   |||
Db        116 ProArgArgAlaAspHisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGlu 135

Qy        811 ACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870
              ||||||||||||:::|||||||||||||||||||||   ||||||      |||   |||
Db        136 ThrProSerLeuLeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaCysLeu***Ser 155

Qy        871 CAGCTACTCAGGAGTCTGAGGCAGGAG 897
              |||||||||      ||||||||||||
Db        156 GlnLeuLeuGlyArgLeuArgGlnGlu 164

RESULT 12
```

```
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:              9.81e-35       Length:        591
Score:                  394.00         Matches:       113
Percent Similarity:     40.9%          Conservative:  6
Best Local Similarity:  38.8%          Mismatches:    39
Query Match:            24.0%          Indels:        133
DB:                     1              Gaps:          4

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU8_HUMAN (1-591)

Qy         31 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 90
              |||   |||       |||||||||   |||||||||          ||| |||   |||
Db          6 GlyAlaValAlaHisAlaCysAsnProSerThrLeu---------Gly-GlyArgSerPr 22

Qy         91 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 150
              ||||||||||||||||||||||||||||       |||   ||||||:::||||||||||
Db         22 oGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThrLysAs 42

Qy        151 TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGA 210
              ||||||||||||          ||||||:::||||||||||||||||||||
Db         42 nThrLysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGl 62

Qy        211 GGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACT 270
              ||||   |||   ||||||||||||||:::|||||||||   |||||||||   |||||
Db         62 uAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLe 82

Qy        271 GCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 330
              ||||||||||||||||||   |||||||||   |||||||||
Db         82 uHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys*********AlaGlyGl 102

Qy        331 AGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTA 390
              |||       |||
Db        102 yGlyArgAlaArg---------------------------------------------- 106

Qy        391 TATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAA 450
                                                                  ||||||
Db        107 ----------------------------------------------------TrpLeu----- 108

Qy        451 GTAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTG 510

Db        108 ------------------------------------------------------------ 108

Qy        511 GAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTG 570

Db        108 ------------------------------------------------------------ 108

Qy        571 GGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTT 630

Db        108 ------------------------------------------------------------ 108

Qy        631 CTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTT 690

Db        108 ------------------------------------------------------------ 108

Qy        691 AAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGG 750
                                          :::   ::: ||||||||||||||||||
Db        109 ------------------------------ThrProVal-IleProAlaLeuTrpGlu- 117

Qy        751 CCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAA 810
                     ||||||||||||||||||||   ||||||||||||||||||   |||
Db        118 --------Glu-AspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGlu 134

Qy        811 ACACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCC 870

                ||||||||||||:::||||||||||||||||||||   ||||||      |||   |||
Db        135 ThrProSerLeuLeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***Ser 154

Qy        871 CAGCTACTCAGGAGTCTGAGGCAGGAG 897
                  |||||||||      ||||||||||
Db        155 GlnLeuLeuGlyArgLeuArgGlnGlu 163

RESULT 13
Q6ZSI5_HUMAN
ID   Q6ZSI5_HUMAN PRELIMINARY;      PRT;   179 AA.
AC   Q6ZSI5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ45499.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Thalamus;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK127408; BAC86963.1; -; mRNA.
SQ   SEQUENCE   179 AA;  19118 MW;  52286CCCA5A4458E CRC64;

Alignment Scores:
Pred. No.:              1.43e-34       Length:        179
Score:                  391.50         Matches:       104
Percent Similarity:     41.4%          Conservative:  14
Best Local Similarity:  36.5%          Mismatches:    33
Query Match:            23.9%          Indels:        134
DB:                     2              Gaps:          5

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x Q6ZSI5_HUMAN (1-179)

Qy        899 TACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAGCT 840
              :::|||||||||   |||  ||||||||||||||||||      ||||||||||||||
Db         29 PheSerCysLeuSerLeuGlnSerSerTrpAspTyrArgHisAlaProProCysProAla 48

Qy        839 AATTTTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTCAAACT 780
              ||||||:::   :::       ||| ||||||||||||   ||||||   ||||||||
Db         49 GlyPheCysLeuCysValGluThrGly-PheHisHisValGlyGlnAlaSerLeuLysLe 68

Qy        779 CCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGAGC 720
              |   ||||||||||||||   |||   |||||||||   |||||||||
Db         68 uLeuThrSerSerAspProProAlaLeuAlaSerGlnIleAlaGlyIle---------- 84

Qy        719 CACCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATC 660

Db         84 ------------------------------------------------------------ 84

Qy        659 CTTAAATGGAATACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTT 600
                                                  |||||| :::|||
Db         85 ---------------------------------ThrGlyMet-IleHis---------- 89

Qy        599 TCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTAT 540

Db         89 ------------------------------------------------------------ 89

Qy        539 CTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTT 480

Db         89 ------------------------------------------------------------ 89

Qy        479 TGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACA 420
```

```
                  |||
Db         90 --------His---------------------------------------------------- 90

Qy        419 GTTAAAATGCATTTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAA 360

Db         90 ------------------------------------------------------------ 90

Qy        359 TCTTATTAGTTTTGTTTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGA 300
                              :::   ||||||           :::||||||||||
Db         91 --------------------AlaGlnProLeu-----------LeuPheLeuArgGlnG 100

Qy        299 GTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCT 240
                 :::   :::   ||||||||||||||||||      |||   |||||||||   ||||
Db        100 lyValThrLeuLeuProArgLeuGluCysSerSerAlaIleLeuAlaHisCysIleLeuC 120

Qy        239 GCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCA 180
              ||||||||||||||||||:::|||   ||||||||||||:::|||||||||||||||||:
Db        120 ysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyV 140

Qy        179 TGCATCACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTG 120
              ::                ||||||||||||||||:::||||||:::||||||||||||:::
Db        140 al---------CysProLeuIlePheValPheValValGluMetGlyPheHisHisVal- 156

Qy        119 ACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTA 60
                 |||||||||   |||   |||||||||||||||||||||||||      ||||||
Db        157 AlaArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProLysVal 176

Qy         59 CTGTGGTTA 51
              |||   |||
Db        177 LeuArgLeu 179
```

RESULT 14
O60448_HUMAN
ID   O60448_HUMAN PRELIMINARY;   PRT;   375 AA.
AC   O60448;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Neuronal thread protein AD7c-NTP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RX   MEDLINE=98064067; PubMed=9399956;
RA   Monte S.M., Ghanbari K., Frey W.H., Beheshti I., Averback P.,
RA   Hauser S.L., Ghanbari H.A., Wands J.R.;
RT   "Characterization of the AD7C-NTP cDNA expression in Alzheimer's
RT   disease and measurement of a 41-kD protein in cerebrospinal fluid.";
RL   J. Clin. Invest. 100:3093-3104(1997).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RA   de la Monte S.M., Wands J.R.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF010144; AAC08737.1; -; mRNA.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0006915; P:apoptosis; TAS.
DR   GO; GO:0007417; P:central nervous system development; TAS.
SQ   SEQUENCE   375 AA; 41720 MW; 955443950A5BFEFD CRC64;

Alignment Scores:

Pred. No.:	2.84e-34	Length:	375
Score:	389.50	Matches:	128
Percent Similarity:	46.2%	Conservative:	40
Best Local Similarity:	35.2%	Mismatches:	88
Query Match:	23.8%	Indels:	108
DB:	2	Gaps:	15

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x O60448_HUMAN (1-375)

```
Qy        901 GATACTCCTGCCTCAGACTCCTGAGTAGCTGGGACTACAGGTGCTTGCCACCATGCCCAG 842
              |||:::|||||||||   |||   |||||||||   ||||||   |||   ||||||:::
Db         29 AspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCysThrHisAlaArg 48

Qy        841 CTAATT---TTTGTATTTTTATTAGAGAGGGTGTTTCATCATGTTGTCCAGGCTGGTCTC 785
              ||||||   :::   ||||||:::|||       |||   ||||||   ||||||||||||
Db         49 LeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGlyGlnAlaGlyLeu 68

Qy        784 AAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCG 725
              :::|||   ||||||   ||||||       ||||||||||||||||||    ||||||
Db         69 GluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla-ArgTyrArg-T 88

Qy        724 TGAGCCACCATGCCCAA---------------------------------------------- 708
                  |||||||||:::
Db         88 hrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArgValSerLeuM 108

Qy        707 -----CCTAGTTTCTCT---ACTTTAAAGCAT-----------GTATTTTAAAAATTAT 668
                  ||||||:::|||      ||||||                        |||   |
Db        108 etCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeuProLysCysT 128

Qy        667 GGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAA------------------- 627
              ||:::                  :::|||:::
Db        128 rpAsp-------------------------TyrArgArgAlaAlaValProGlyLeuP 139

Qy        626 --------------ATTAGGTATGCCTGC---ACATGGAAATTTTCCCAAATAAAA---- 588
                           :::|||:::   |||   |||            :::::::::
Db        139 heIleLeuPhePheLeuArgHisArgCysProThrLeuThrGlnAspGluValGlnTrpC 159

Qy        587 -----AAAGCTAGTGTATGCCCCAAAAATCTGGAG----------------------G 557
                    :::|||:::   |||          |||
Db        159 ysAspHisSerSerLeuGlnProSerThrProGluIleLysHisProProAlaSerAlaS 179

Qy        556 AATATGTAACTGGT-----------------TATCTTTGGGTAGTAGGATTATAGATGC 515
              |||   |||                   |||   |||::::::   :::
Db        179 erGlnValAlaGlyThrLysAspMetHisHisTyrThrTrpLeuIlePheIle------- 196

Qy        514 CTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT---------- 465
                 |||          :::|||   :::      :::   :::|||
Db        197 --------PheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyV 214

Qy        464 ----------------------------------------TTTGGCTTGTTTACT- 450
                                                     |||   ||||||:::
Db        214 alGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerC 234

Qy        449 ------------------------------------------------------TTTA 446
                                                                     |||:
Db        234 ysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheV 254

Qy        445 TAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACA------T 392
              ::      :::          |||:::         |||      |||:::
Db        254 alPheLeuValGluMetGlyPheThrMetPheAlaArgLeuIleLeuIleSerGlyProC 274

Qy        391 ATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGTTTCTCTCCGATCC 332
              :::|||      ||||||                |||:::   ||||||    :::
Db        274 ysAspLeuProAlaSerAlaSerGlnSerAlaGly-IleThrGlyValSerHisHisAla 293

Qy        331 TTTATTTTATTTATTTAT---TTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGA 275
              ::::::|||   :::      ||||||:::||||||      |||   |||||||||
Db        294 ArgLeuIlePheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGly 313

Qy        274 GTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCT 215
              |||||||||   :::|||||||||||||||||||||||||||||      |||:::||||||
Db        314 ValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLysArgPheSer 333

Qy        214 TGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTT 155
              |||||||||||||||||||||||||||||||||   |||          ||||||||||||||||
Db        334 CysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisProAlaAsnPhe 353

Qy        154 TGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCGAACTTCTCAC 95
              |||||||||   |||      ||||||||||||      |||||||||:::|||
Db        354 CysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerGlnThrProAsp 373
```

```
Qy          94 CTCAAG 89
               |||:::
Db         374 LeuArg 375

RESULT 15
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:              3.33e-32      Length:          587
Score:                  371.50        Matches:         101
Percent Similarity:     40.2%         Conservative:    16
Best Local Similarity:  34.7%         Mismatches:      43
Query Match:            22.7%         Indels:          131
DB:                     1             Gaps:            1

US-10-607-806-1-T9182_COPY_8700_9600 (1-901) x ALU2_HUMAN (1-587)

Qy          32 GCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTT 91
               |||:::   |||   ||||||         ||||||:::::::::   |||||||||
Db           3 AlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySerArg 22

Qy          92 GAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAAT 151
                  |||::::::   ||||||   |||   |||   |||||||||   ||||||   ||||
Db          23 --GlyGlnGluIleGluThrIleLeuAlaAsnThrValLysProArgLeuTyr***LysT 42

Qy         152 ACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAG 211
               ||||||||      |||||||||      ||||||||||||||||||||||||||   |
Db          42 yrLysAsn***ProGlyValValAlaGlyAlaCysSerProSerTyrSerGlyGly***G 62

Qy         212 GCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTG 271
               ||:::|||:::|||   |||   |||||||||:::|||||||||||||||      ||||
Db          62 lyArgArgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrA 82

Qy         272 CACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAA 331
               |||||||||||||||||   |||||||||   |||||||||                 |
Db          82 laLeuGlnProGlyArgGlnSerGluThrProSerGlnLysLys*********AlaGlyA 102

Qy         332 GGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTAT 391
               ||   |||   :::
Db         102 rgGlyGlySerArgLeu---------------------------------------------- 107

Qy         392 ATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAG 451

Db         107 ---------------------------------------------------------------- 107

Qy         452 TAAACAAGCCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGG 511

Db         107 ---------------------------------------------------------------- 107

Qy         512 AAGGCATCTATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGG 571

Db         107 ---------------------------------------------------------------- 107

Qy         572 GGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTC 631

Db         107 ---------------------------------------------------------------- 107

Qy         632 TAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTA 691

Db         107 ---------------------------------------------------------------- 107

Qy         692 AAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGC 751
                                                  ||||||||||||||||||
Db         108 ------------------------------------***SerGlnHisPheGlyArgP 115
```

```
Qy        752 CGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAACATGATGAAA 811
              |||||:::   ||||||  |||||||||      ||||||:::|||   |||      ||
Db        115 roArgArgAlaAspHis--GluValArgArgSerArgProSerTrpLeuThrArg***As 134

Qy        812 CACCCTCTCTAATAAAAATACAAAAATTAGCTGGGCATGGTGGCAAGCACCTGTAGTCCC 871
              |   :::|||   ||||||||||||||||||   |||||||||:::|||||||||||||||
Db        134 nProValSerThrLysAsnThrLysIleSerArgAlaTrpTrpArgAlaProValValPr 154

Qy        872 AGCTACTCAGGAGTCTGAGGCAGGA 896
              |||||||:::|||:::|||||||||
Db        154 oAlaThrArgGluAlaGluAlaGly 162


Search completed: February  8, 2006, 11:14:23
Job time : 217.737 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:22:47 ; Search time 70.7692 Seconds
(without alignments)
3105.549 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 4426
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092016_27646/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_605_@runat_08022006_092016_27646 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	475	10.7	7067	6	AAE35265	Aae35265 Human P45
c	2	468.5	10.6	7052	6	AAE35266	Aae35266 Human P45
	3	455.5	10.3	7052	6	AAE35266	Aae35266 Human P45
c	4	451	10.2	7036	6	AAE35264	Aae35264 Human P45
	5	438.5	9.9	215	4	AAU31513	Aau31513 Novel hum
	6	427.5	9.7	217	7	ADC87151	Adc87151 Human GPC
	7	420.5	9.5	7067	6	AAE35265	Aae35265 Human P45
	8	416	9.4	673	6	ADA84009	Ada84009 Human POM
	9	413	9.3	185	4	AAO06014	Aao06014 Human pol
	10	405	9.2	232	5	ADK36978	Adk36978 Novel hum
	11	400	9.0	302	5	ADK36936	Adk36936 Novel hum
	12	400	9.0	361	5	ABG68738	Abg68738 Human pro
	13	400	9.0	449	5	ABJ01057	Abj01057 Ovary cel
	14	397.5	9.0	304	4	AAU30391	Aau30391 Novel hum
	15	397.5	9.0	375	3	AAB01399	Aab01399 Neuron-as
	16	397.5	9.0	375	5	ABB81538	Abb81538 Neural th
	17	397.5	9.0	375	5	AAE29142	Aae29142 AD7c-NTP
	18	397.5	9.0	375	6	ABR63268	Abr63268 AD7c-NTP
	19	397.5	9.0	375	6	ADA84017	Ada84017 Human POM
	20	397.5	9.0	375	6	ABU03024	Abu03024 Human neu
	21	397.5	9.0	375	6	ABB99774	Abb99774 Amino aci
	22	397.5	9.0	375	6	ABJ19445	Abj19445 AD7C-neur
	23	397.5	9.0	375	7	ADB37642	Adb37642 Human neu
	24	397.5	9.0	375	8	ADR14409	Adr14409 Human NF-
	25	397.5	9.0	375	9	AEA30020	Aea30020 Human neu
	26	396.5	9.0	239	8	ADQ65417	Adq65417 Novel hum
c	27	391.5	8.8	673	6	ADA84009	Ada84009 Human POM
	28	391.5	8.8	7036	6	AAE35264	Aae35264 Human P45
	29	390	8.8	382	4	AAU32610	Aau32610 Novel hum
	30	390	8.8	382	4	AAU31818	Aau31818 Novel hum
	31	390	8.8	382	4	AAU32707	Aau32707 Novel hum
	32	387.5	8.8	396	4	AAU30455	Aau30455 Novel hum
	33	386	8.7	179	8	ADR09211	Adr09211 Human pro
	34	384.5	8.7	158	7	ADB64893	Adb64893 Human pro
	35	384	8.7	324	4	AAU29573	Aau29573 Novel hum
	36	384	8.7	324	4	AAU29579	Aau29579 Novel hum
	37	384	8.7	397	2	AAR95913	Aar95913 Neural th
	38	383.5	8.7	257	4	AAO10294	Aao10294 Human pol
	39	381	8.6	241	8	ADQ66812	Adq66812 Novel hum
	40	381	8.6	527	7	ADD27133	Add27133 Human adi
	41	380.5	8.6	175	4	AAU31857	Aau31857 Novel hum
	42	380.5	8.6	250	4	AAU31823	Aau31823 Novel hum
	43	380.5	8.6	250	7	ADE09441	Ade09441 Novel pro.
	44	377	8.5	215	4	ABG07921	Abg07921 Novel hum
	45	376	8.5	189	7	ADB65042	Adb65042 Human pro

ALIGNMENTS

RESULT 1
AAE35265
ID AAE35265 standard; protein; 7067 AA.
XX
AC AAE35265;
XX
DT 28-MAY-2003 (first entry)
XX
DE Human P450RAI-3 protein #2.
XX
KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW enzyme.
XX
OS Homo sapiens.
XX
PN WO200295034-A2.
XX
PD 28-NOV-2002.
XX
PF 23-MAY-2002; 2002WO-CA000758.
XX
PR 23-MAY-2001; 2001US-0292531P.
XX
PA (CYTO-) CYTOCHROMA INC.
XX
PI Wisniewski J, Petkovich PM, Ramshaw H;
XX
DR WPI; 2003-112152/10.
DR N-PSDB; AAD53861.
XX
PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT diagnosing, preventing, or treating disorders with aberrant expression or

PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
XX
PS Example 1; Fig 1; 231pp; English.
XX
CC The invention relates to retinoic acid metabolising cytochrome P450,
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX
SQ Sequence 7067 AA;

Alignment Scores:

Pred. No.:	1.66e-45	Length:	7067
Score:	475.00	Matches:	218
Percent Similarity:	39.0%	Conservative:	68
Best Local Similarity:	29.7%	Mismatches:	204
Query Match:	10.7%	Indels:	243
DB:	6	Gaps:	31

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35265 (1-7067)

```
Qy       2497 CCAGGCTGGTCTCAAACTCCGGACCTCAAGTGATCCATTTGCCTCGGCCTCCCAAAGTGC 2438
Db         27 ProGlyTrpSerArgSerHisValIle--MetIleHisProProTrpProProLysValL 46

Qy       2437 TGGGATACAGGCGTGAGCCA---CCATGCCCAACCTAGTTTCTCTACTTTAAAGCATGTA 2381
Db         46 euGlyLeuGlnAlaAlaThrValProGlyPro--------------------AlaAlaL 59

Qy       2380 TTTTAAAAATTATGGAACATCCTTAAATGGAATACTACATAGACATTTAGAAAAATTAGG 2321
Db         59 euPheLeuIleAla---------------------------------------------- 63

Qy       2320 TATGCCTGCACATGGAAATTTTCCCAAATAAAAAAAGCTAGTGTATGCCCCAAAAATCTG 2261
Db         64 ----------------------------LysLysAspLysCys---ProSerThrAsp- 72

Qy       2260 GAGGAATATGTAACTGGTTATCTTTGGGTAGTAGGATTATAGATGCCTTCCACTTTCCAT 2201
Db         73 ----------AspTrpIleAsn------------------------------------- 76

Qy       2200 AATAGATATTTATGTAGAGTTTGAAGTTGCATTACTTTTGGCTTGTTTACTTTTATAACC 2141
Db         77 ----------------------LysValTrpTyrIleHisThrIleTyrTyrTyrLeuS 89

Qy       2140 AAAATAAACAATAAGATTACAGTTAAAATGCATTTAAACTTTATTACATATAATTTACGT 2081
Db         89 erIleLysArgAsnGluValIleMetHisAlaThrThrThrLeuLysIle---------- 105

Qy       2080 ATCTCTGCTTACCTTCAGGA----------------------ATCTTATTAGTTTTGT 2045
Db        106 --LeuCysLeuSerGluArgSerGlnLeuGlnLysThrArgValPheLeuPheCysLeuP 125

Qy       2044 TTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGC 1985
Db        125 he------------CysPheValLeuPheCysPheLysArgGlnGlyLeuAlaLeuLeuP 141

Qy       1984 CCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1925
Db        141 roIleLeuGlu-Tyr------SerSerLeuLeuThrAlaAlaSerThrPheTrpThrGln 158

Qy       1924 GCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCC 1865
Db        159 AlaSerLeuProProGlnProProLys---LeuAlaLeuGlnValCysThrThrThrAla 177

Qy       1864 AGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCG 1805
Db        178 Ser---PheLeu-IleValCysArgAsnArgValSerProCysCysProGlyTrpSerGl 196

Qy       1804 AACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCT 1745
Db        196 nThrProSerLeuLys---------------------------------------------- 201

Qy       1744 TGAGTCACTGTGCCCAACTGAGTTATTTTTAATTAATATGAACAAACATCCTCTAATGCA 1685
Db        202 ------------ProTrpProLeuLysValLeuGlyLeuGlnGlyAlaThrMetValGl 217

Qy       1684 GGTTTCTTGTCTTAGTTGCAAACATTAAAAAAAAACTGTGTGTGAGTGTGTGTGTGTGTG 1625
Db        217 y--------------------------GlnLysThrThrCysTyrLeuIleProIle-- 227

Qy       1624 TGTATTTCTCTGGGGA----------------------------GAGGGTTCATAGCT 1595
Db        228 ----------TrpGlyValIleAlaAsnGlyTyrGlyValSerPheArgGlySerLysIl 244

Qy       1594 TTCATCAGATTTGTAAATTTGGCCTTCTAAAGTGGAGCAGGAGGATGTATAAGCAAATTA 1535
Db        244 eAspAspGly-CysThr------IleLeuIleGluSerLysProLeuIleCysAlaLeuM 262

Qy       1534 AGCCATATTATTCAGTGACTAAAATAGTAACAATGAAAAAAC------------AGTAAC 1487
Db        262 etSerLeuTyrGlyThrSerGlnSerCysLeuLys-LysAsnPheSerTyrIleThrArg 281

Qy       1486 AATAAAGGTTTGAGGCAATATGAAAAAGTCATAACAAGGTGATAAGTGAAAAAGCAGAAT 1427
Db        282 AsnArgGlyLeu---------------------------------------------- 285

Qy       1426 GCAAGATTGTATTTATTCTGCAAAGTTGTCAAGATTGTTTTAAAAAAAGAAAAAGAAAGT 1367
Db        286 --------------------------------IleIleLeuLysLeuLysGlnLysThr 294

Qy       1366 TTATATCTAGGCTGT----------------------GCATTCATCAATTATTTATTG 1331
Db        295 LysMetValGlyLeuProLysLysArgAlaGlyIleLysGlyLysPheSerTyrValIle 314

Qy       1330 AGC------ACCTACTATATGTCAGACACTGCTCTAGATCTGGGGATATAGCAATAAAGA 1277
Db        315 AlaValLysThrGlnAsnLysAsnLysProIleArgAspSerGluIleIleThrArgThr 334

Qy       1276 AAATAGACAGTAATCGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAG 1217
Db        335 LysArgGlyValLysGlyGlnThrGlnTrpLeuThrProValIleProAlaPheTrpGlu 354

Qy       1216 ACTGAGGCAGGCAGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAA 1157
Db        355 AlaGluValGlyGlySerLeuGluAlaArgSerLeuArgProAlaTrpAlaThrTrp--- 373

Qy       1156 AACCCCATCTCTGCTAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAAT 1097
Db        374 AsnProValSerThrLysAsnThr-LysIleSerArgAlaTrpTrpTyrValProValVa 393

Qy       1096 TCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCA 1037
Db        393 lProAlaAlaTrpGluAlaGluAlaGlnGluLeuLeuGluPro----------------- 407

Qy       1036 GTGAGCTGAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 977
Db        407 ---------------------------------------------------------- 407

Qy        976 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 917
Db        407 ---------------------------------------------------------- 407

Qy        916 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 857
Db        408 -------GluAlaArgCysGlyGlySerArgLeu---SerGlnHisPheGlyArgProAr 424

Qy        856 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCC------TGGCCAACGTAGTAA 803
Db        424 gTrpValAsp--HisGluValArgSerLeuArgProAlaProThrTrp------------ 439
```

```
Qy        802 GACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAG 743
              :::|||:::|||      |||:::      ||||||      ||||||
Db        440 AsnProIleSer---AsnLysTyrLysLysLeuAlaGlyHisGlyGlyVal---CysLeu 457

Qy        742 TCCCAGCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGC 683
              |||||||||   |||||||||      :::   |||:::   ||||||   ||||||
Db        458 CysGlnLeuLeuArgArgLeuArgGlnGluAsnCysLeuAsnLeuGlyGlyGlyGlyCys 477

Qy        682 AGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCT 623
              |||:::            ||||||          |||   |||::::::|||   ||||||
Db        478 SerGluProArgSerArgHisCysThrProThrTrpAlaThrLysArgAspSerValSer 497

Qy        622 AAACAACAACAAAAAAAAACCCAAACTGTCACCACTTACAAGCTGTGTGATCTTGGGCAA 563
              |||:::::::::||||||   :::   :::   :::   :::               :::
Db        498 LysLysLysLysLysLysGluLeuLeuGluSerGlyArgArgArgLeuGlnAlaGlu 517

Qy        562 ATAGCTCAGCCTTCTGAGCCTCAGTTTTCTTATCTGTTAAGTGAGGATAATTGTACCTAT 503
              ||||||            |||      |||   |||
Db        518 IleAla------------ProLeuHisSerSerLeuGlyAspArgAlaArgLeuCys--- 532

Qy        502 CCCATAGTGTTGTTCTGAGGATTAAAGGAGACACTGCCCAGAACCAAGAAAATTAACACT 443
                                 ||||||   |||                |||
Db        533 --------------------LeuLysThrThrThrThrGluAsnThrLysAsnLysIle 545

Qy        442 AAATCA---TGGCTGTTGTTA 425
              :::|||   ||||||||||||
Db        546 ArgSerThrTrpLeuLeuLeu 552
```

RESULT 2
AAE35266
ID AAE35266 standard; protein; 7052 AA.
XX
AC AAE35266;
XX
DT 28-MAY-2003 (first entry)
XX
DE Human P450RAI-3 protein #3.
XX
KW Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW enzyme.
XX
OS Homo sapiens.
XX
PN WO200295034-A2.
XX
PD 28-NOV-2002.
XX
PF 23-MAY-2002; 2002WO-CA000758.
XX
PR 23-MAY-2001; 2001US-0292531P.
XX
PA (CYTO-) CYTOCHROMA INC.
XX
PI Wisniewski J, Petkovich PM, Ramshaw H;
XX
DR WPI; 2003-112152/10.
DR N-PSDB; AAD53861.
XX
PT New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT diagnosing, preventing, or treating disorders with aberrant expression or
PT activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT psoriasis.
XX
PS Example 1; Fig 1; 231pp; English.
XX
CC The invention relates to retinoic acid metabolising cytochrome P450,
CC P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC invention are useful for diagnosing, preventing, ameliorating and/or
CC treating disorders associated with the aberrant expression or activity of
CC the P450RAI such as diseases related to vitamin A and retinoic acid
CC metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC present sequence is human P450RAI protein
XX
SQ Sequence 7052 AA;

Alignment Scores:

Pred. No.:	9.78e-45	Length:	7052
Score:	468.50	Matches:	199
Percent Similarity:	37.7%	Conservative:	62
Best Local Similarity:	28.7%	Mismatches:	170
Query Match:	10.6%	Indels:	264
DB:	6	Gaps:	28

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35266 (1-7052)

```
Qy       2464 TCCATTTGCCTCGGCCTCCCAAAGTGCTGGGATACAGGCGTGAGCCACCATGCCCAACCT 2405
              |||      ||||||||||||||||||||||||            :::      :::|||
Db         39 SerThrLeuLeuGlyLeuProLysCysTrpAsp-----------TyrArgArgGluPro 54

Qy       2404 AGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATCCTTAAATGGAATACT 2345
                              :::|||   |||   |||      |||:::
Db         55 ArgCysProAlaGlnGlnHisTyrPheProLysLysIleAsnVal-------------- 69

Qy       2344 ACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTTTCCCAAATAAAAAAA 2285
                                                                 |||:::
Db         70 ----------------------------------------------HisGlnLeuMetIle 74

Qy       2284 GCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTATCTTTGGGTAGTAGGA 2225
                :::   |||      :::::::::      :::
Db         75 GlyThrArgCysGlyThrSerIleGlnTyrIleIle---------------------- 86

Qy       2224 TTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATTACT 2165
                                                :::|||:::
Db         87 --------------------------------IleCysGlnLys-------------- 90

Qy       2164 TTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGTTAAAATGCATTTA 2105
                |||:::      :::  :::      |||
Db         91 ---GlyMetLysCysMetLeuGlnArgAspGluProLysTyr---------------- 103

Qy       2104 AACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTTTGT 2045
                     |||   ::::::   ::::::|||   :::         ||||||
Db        104 ------------TyrAlaValLysGluAlaSerTyrLysArgProGlyArgSerPheCys 119

Qy       2044 TTCTCTCCGATCCTTTATTTTATTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGC 1985
              |||         :::         ::::::          ||||||   ||||||
Db        120 PheValCysPheValLeuCysCysPheValLeuAsnArgGlyArgValLeuLeuCysCys 139

Qy       1984 CCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAA 1925
              |||   ||||||   |||              ||||||||||||   |||         ||
Db        140 ProSerTrpSerThrVal------His-CysSerLeuGlnProProProSerGlyLeuLy 157

Qy       1924 GCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATCACTACACCC 1865
              |      |||   ||||||||||||||||      ||||||   |||         ||
Db        157 sLeuAlaSerHisLeuSerLeuProSerSer---HisTyrArgCysAlaProProGlnPr 176

Qy       1864 AGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGGCTGGTCTCG 1805
              |||||||||   :::::: ||||||   |||   |||   :::   ||||||||||||:
Db        176 oAlaAsnPhe---LeuTyr-ValGluThrGlySerHisArgValAlaGlnAlaGlyLeuL 195

Qy       1804 AACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGGTTACAGGCT 1745
              ::||||||   ||||||                                        ||||
Db        195 ysLeuLeuValSerSer-----------------------------------LeuGlyL 203

Qy       1744 TGAGTCACTGTGCCCAACTGAGTTATTTTTAATTAATATGAACAAACATCCTCTAATGCA 1685
              |||||   |||
Db        203 euSerLysCysAsp------------------------------------------- 207

Qy       1684 GGTTTCTTGTCTTAGTTGCAAACATTAAAAAAAAACTGTGTGTGAGTGTGTGTGTGTG 1625

Db        207 ------------------------------------------------------------ 207

Qy       1624 TGTATTTCTCTGGGGAGAGGGTTCATAGCTTTCATCAGATTTGTAAATTTGGCCTTCTAA 1565
                  :::   ||||||
Db        208 -----TyrArgGlyGluPro-------------------------------------P 213
```

```
Qy       1564 AGTGGAGCAGGAGGATGTATAAGCAAATTAAGCCATAT---------------------- 1527
              |||:::||||||   :::      |||   |||:::
Db        213 roTrpAlaArgArgProHis---ValIlePheProPheGlyGluLeuLeuMetAlaThrG 232

Qy       1526 --------TATTCAGTGACTAAAATAGTAACAATGAAAAAACAGTAACAATAAAGGTTTG 1475
                           |||::::::   |||      |||
Db        232 lyPheLeuSerGlyAlaLeuLysLeuIleGluMetValAlaGln---------------- 246

Qy       1474 AGGCAATATGAAAAAGTCATAACAAGGTGATAAGTGAAAAAGCAGAATGC---------- 1425
                    :::|||   |||      :::      |||      |||
Db        247 -----PheCysGluSerGlnAsnHisPheValProPheLysValAsnCysMetValHisL 265

Qy       1424 --------------AAGATTGTATTTATTCTGCAAAGTTGTCAAGATTGTTTT------- 1386
                           ||||||      ::::::|||      :::      :::
Db        265 euLysAlaValLysLysIlePheLysValIleGlnGluIleGluValLeuTyrSerTyrL 285

Qy       1385 -----AAAAAAAGA--------------------AAAAGAAAGTTTATATCTAGGCTGT 1352
                  |||   |||                     :::|||:::   :::|||::::::
Db        285 ysAsnLysAsnArgGlnLysTrpAspSerTyrProArgArgGluLeuValSerLysValA 305

Qy       1351 GCATTCATCAATTATTTATTGAGCACCTACTATATGTCAGACACTGCTCTAGATCTGGGG 1292
                  |||   |||          :::   |||   :::|||       |||      |||
Db        305 snSerHisMetLeu------AsnAsnLysLeuLysIleArgLysIleSerLeuLeuGlyI 323

Qy       1291 ATATAGCAATAAAGAAAATAGACAGTAA----TCGGCCAGGCACAGTGACTCACGCCTAT 1236
                :::          ::::::                  ||||||   |||   |||   |||
Db        323 leValArgSerLeuGluLeuLysGluGluLeuLysAlaArgArgSerGlySerHisLeu- 342

Qy       1235 AATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAGGAGTTTGAGACC 1176
                  |||||||||||||||   |||      ||||||||||||   ||||||   ||||
Db        343 --SerGlnHisPheGlyArgProArgTrpValAspHisLeuArgProGlyVal---AspG 361

Qy       1175 AGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAATTAGCCAAGCAT 1116
                |||||||||||||||:::|||||||||||   |||   ||| |||||||||
Db        361 lnProGlyGlnHisGlyGluThrProSerLeuGlnLysThrGln-LysLeuAlaGlyArg 380

Qy       1115 GGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAAC 1056
                ||||||      |||      |||||||||   |||||||||   :::|||   ||||||
Db        381 GlyGlyMetCysLeu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnCysLeuAsn 399

Qy       1055 ------CCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGGTTGAGGTGATGAAGTGAGA 1002
                     ||||||
Db        400 LeuArgProGlyAla---------------------------------------------- 404

Qy       1001 CTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACA 942

Db        404 ------------------------------------------------------------ 404

Qy        941 TTCTATTGGAAAGATGGAATGTAAAGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGT 882
                                                          |||||||||   |||
Db        405 ---------------------------------------------ValAlaHisAlaCys 409

Qy        881 AATCCCAGCACTTTGGGATGCCAAGGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACA 822
                |||||||||||||||||   |||    |||   :::    :::   ||||||   |||:
Db        410 AsnProSerThrLeuGlyGlyGlnAsp-GlyTrpIle-ThrArgSerGlyVal---AspG 428

Qy        821 AGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCG 762
                ::|||   |||      :::|||   ||||||   ||||||      |||   ||||||:
Db        428 lnProAspGlnHisGlyGluThrProSerLeuThrLysAsnThrLysAsn---ProGlyM 447

Qy        761 TGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCTG---------------------- 722
                ::|||      |||   :::   ||||||   ||||||
Db        447 etValAlaCysAlaCysAsnAlaSerTyrSerGlyGlyGlyArgArgThrAlaThrTrpG 467

Qy        721 ------------------------------------------------------AGGTCAG 715
                                                                     :::|||
Db        467 luValGluValAlaValSerArgAspHisAlaThrAlaLeuGlnProGlyGlnGlnSerG 487

Qy        714 CAGATCCC--------------------------------TTGAGCTCAGGAGGTTGAG 688
                  |||                                 |||:::   ||||||   |
Db        487 luThrProSerGlnLysLysLysLysLysLysAsnCysLeuAsnProGlyGlyGluG 507
```

```
Qy        687 GCTGCAGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTG 628
                |||||||:::      ||||||||||||         ||||||||||||:::|||   |
Db        507 lyCysSerGluProArgLeuHisHisCysThrProValTrpValThrGluArgAspSerV 527

Qy        627 TCTCTAAACAACAACAAAAAAAAAACCCAA 599
                ||||||||||||||||:::|||||||||
Db        527 alSerLysGlnGlnGlnGlnLysThrGln 536

RESULT 3
AAE35266
ID   AAE35266 standard; protein; 7052 AA.
XX
AC   AAE35266;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #3.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7052 AA;
```

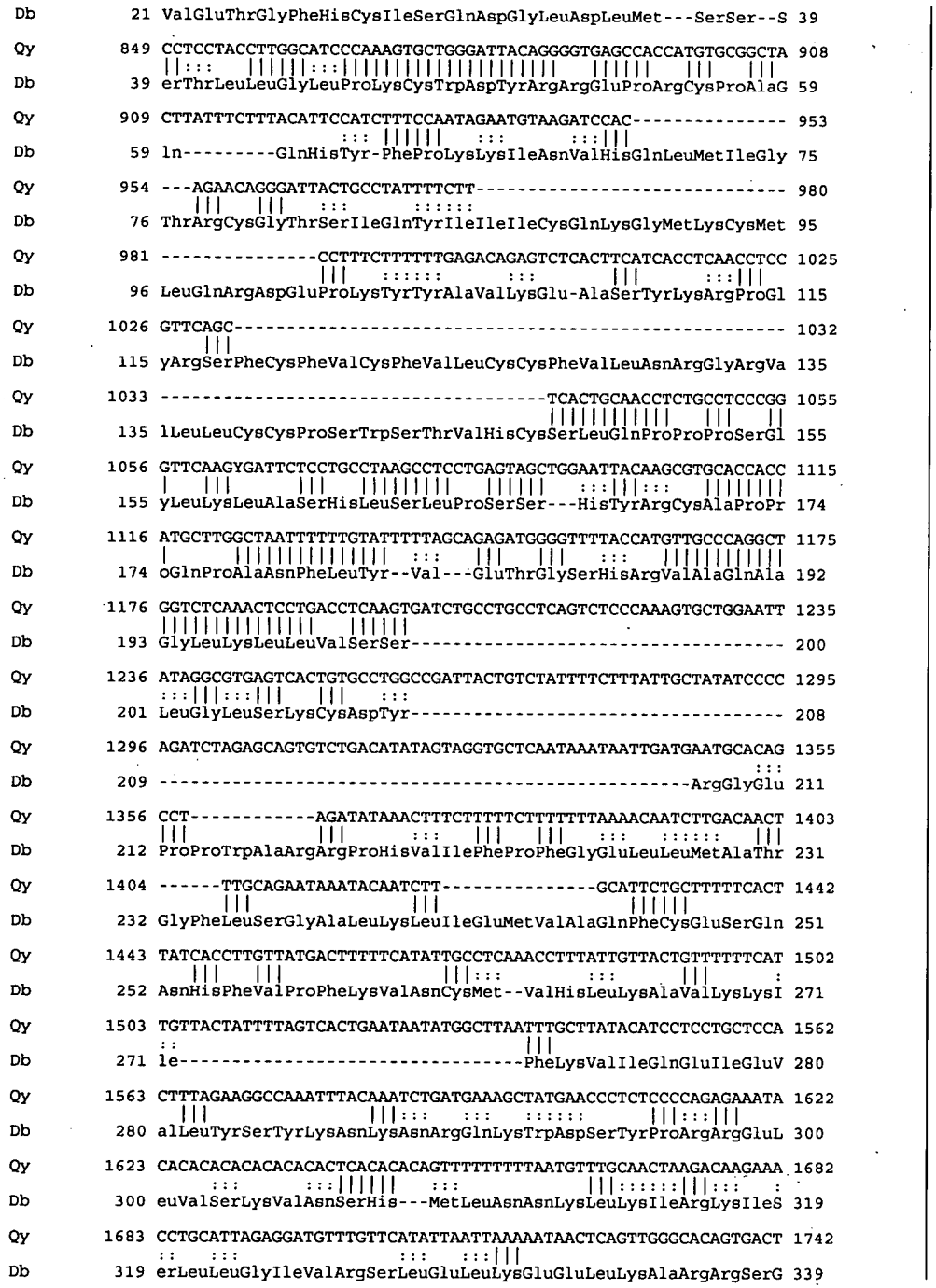
Alignment Scores:

Pred. No.:	3.38e-43	Length:	7052
Score:	455.50	Matches:	208
Percent Similarity:	35.1%	Conservative:	75
Best Local Similarity:	25.8%	Mismatches:	211
Query Match:	10.3%	Indels:	312
DB:	6	Gaps:	29

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35266 (1-7052)

```
Qy        732 AGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT---TTTTTT 788
                |||||||||||||||          ||||||      ||||||              |||
Db          1 SerSerTrpAspTyrArgCysLeuProProCysProAlaAsnPheLeuTyrIlePheLeu 20

Qy        789 GTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAAT 848
                ||||||||||||   :::   :::   |||      |||   |||:::   ||||||  |
```


Db 21 ValGluThrGlyPheHisCysIleSerGlnAspGlyLeuAspLeuMet---SerSer--S 39
Qy 849 CCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTA 908
Db 39 erThrLeuLeuGlyLeuProLysCysTrpAspTyrArgArgGluProArgCysProAlaG 59
Qy 909 CTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC---------- 953
Db 59 ln---------GlnHisTyr-PheProLysLysIleAsnValHisGlnLeuMetIleGly 75
Qy 954 ---AGAACAGGGATTACTGCCTATTTTCTT---------- 980
Db 76 ThrArgCysGlyThrSerIleGlnTyrIleIleIleCysGlnLysGlyMetLysCysMet 95
Qy 981 ----------CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCC 1025
Db 96 LeuGlnArgAspGluProLysTyrTyrAlaValLysGlu-AlaSerTyrLysArgProGl 115
Qy 1026 GTTCAGC---------- 1032
Db 115 yArgSerPheCysPheValCysPheValLeuCysCysPheValLeuAsnArgGlyArgVa 135
Qy 1033 ----------TCACTGCAACCTCTGCCTCCCGG 1055
Db 135 lLeuLeuCysCysProSerTrpSerThrValHisCysSerLeuGlnProProProSerGl 155
Qy 1056 GTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACC 1115
Db 155 yLeuLysLeuAlaSerHisLeuSerLeuProSerSer---HisTyrArgCysAlaProPr 174
Qy 1116 ATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCT 1175
Db 174 oGlnProAlaAsnPheLeuTyr--Val---GluThrGlySerHisArgValAlaGlnAla 192
Qy 1176 GGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATT 1235
Db 193 GlyLeuLysLeuLeuValSerSer---------- 200
Qy 1236 ATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCC 1295
Db 201 LeuGlyLeuSerLysCysAspTyr---------- 208
Qy 1296 AGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAG 1355
Db 209 ----------ArgGlyGlu 211
Qy 1356 CCT----------AGATATAAACTTTCTTTTTCTTTTTTAAAACAATCTTGACAACT 1403
Db 212 ProProTrpAlaArgArgProHisValIlePheProPheGlyGluLeuLeuMetAlaThr 231
Qy 1404 ------TTGCAGAATAAATACAATCTT----------GCATTCTGCTTTTTCACT 1442
Db 232 GlyPheLeuSerGlyAlaLeuLysLeuIleGluMetValAlaGlnPheCysGluSerGln 251
Qy 1443 TATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCAT 1502
Db 252 AsnHisPheValProPheLysValAsnCysMet--ValHisLeuLysAlaValLysLysI 271
Qy 1503 TGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCA 1562
Db 271 le----------PheLysValIleGlnGluIleGluV 280
Qy 1563 CTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATA 1622
Db 280 alLeuTyrSerTyrLysAsnLysAsnArgGlnLysTrpAspSerTyrProArgArgGluL 300
Qy 1623 CACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAA 1682
Db 300 euValSerLysValAsnSerHis---MetLeuAsnAsnLysLeuLysIleArgLysIleS 319
Qy 1683 CCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACT 1742
Db 319 erLeuLeuGlyIleValArgSerLeuGluLeuLysGluGluLeuLysAlaArgArgSerG 339
Qy 1743 CAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAG 1802
Db 339 lySerHisLeuSerGlnHisPheGlyArgProArgTrpValAspHisLeuArgProGlyV 359
Qy 1803 TTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAG 1862
Db 359 al---AspGlnProGlyGlnHisGlyGluThrProSerLeuGlnLysThrGlnLysLeuA 378
Qy 1863 CTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATT 1922
Db 378 laGlyArgGlyGlyMetCysLeu---SerGlnLeuLeuGlyArgLeuArgHisLysAsnC 397
Qy 1923 GCTTGAACCTG---------- 1933
Db 397 ysLeuAsnLeuArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyG 417
Qy 1933 ---------- 1933
Db 417 lnAspGlyTrpIleThrArgSerGlyValAspGlnProAspGlnHisGlyGluThrProS 437
Qy 1933 ---------- 1933
Db 437 erLeuThrLysAsnThrLysAsnProGlyMetValAlaCysAlaCysAsnAlaSerTyrS 457
Qy 1933 ---------- 1933
Db 457 erGlyGlyGlyArgArgThrAlaThrTrpGluValGluValAlaValSerArgAspHisA 477
Qy 1933 ---------- 1933
Db 477 laThrAlaLeuGlnProGlyGlnGlnSerGluThrProSerGlnLysLysLysLysLysL 497
Qy 1934 ----------GGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
Db 497 ysLysAsnCysLeuAsnProGlyGlyGluGlyCysSerGluProArgLeuHisHisCysT 517
Qy 1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGG 2033
Db 517 hrProValTrpValThrGluArgAspSerValSerLysGln------GlnGlnGlnLysT 535
Qy 2034 ATCGGAGAGAAACAAAACTAATAAGA----------TTCC 2063
Db 535 hrGlnLysThrAsnLysGlyValLysHisGlyTyrPheValArgGlnThrValGlyPheM 555
Qy 2064 TGAAGGTAAGCAGAGATACGTAAATTATATGTAATA---------- 2099
Db 555 etIleCysSer-IleLeuLeuThrMetTyrLeuIleLysLeuIleLeuIleAsnProHis 574
Qy 2100 AAGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAG 2159
Db 575 LysPheThrCysMetTyrLeuTyrAlaIleLeuAlaMetValValThrLysLysLysLys 594
Qy 2160 CCAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATC 2219
Db 595 IleLysIleLysGln----------IleTrpMet---------Cys 603
Qy 2220 TATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACA 2279
Db 604 ValAsnProIleMet---------IleLeuLeuTyrSerSerPhePhePhe---------- 617
Qy 2280 CTAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTC 2339
Db 618 ----------PheLysPheGlnThrHisLeuGluValIleAspTyr 629
Qy 2340 TATGTAGTATTC---------CATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTT 2390
Db 630 TyrIleSerLeuAlaLysAsnThrLeuArgThrPheTyr----------IleCysMet 645
Qy 2391 AAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCTGTA---------- 2432
Db 646 TyrSerSerProGluSerTyrLeuLysArgLeuIleProLeuSerSerPheCysSerMet 665

```
Qy       2433 ------------------------------------TCCCAGCACTTTGGGAGGCCG 2453
                                                    |||:::     |||
Db        666 LysCysSerGluLeuThrProSerProHisArgGlySerGluArgIleGlyGluGln 685

Qy       2454 AGGCAAATGGATCAC--------------TTGAGGTCCGGAGTTTGAGACCAGCCTGGA 2498
              ||||||   |||             |||     |||     |||     |||
Db        686 ValGlnMetGlyHisTrpLysArgAlaGlyLeuGlyGluGlyAlaSerAspProGluGly 705

RESULT 4
AAE35264
ID   AAE35264 standard; protein; 7036 AA.
XX
AC   AAE35264;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #1.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7036 AA;

Alignment Scores:
Pred. No.:              1.15e-42        Length:         7036
Score:                  451.00          Matches:        238
Percent Similarity:     38.6%           Conservative:   107
Best Local Similarity:  26.7%           Mismatches:     238
Query Match:            10.2%           Indels:         312
DB:                     6               Gaps:           39

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35264 (1-7036)

Qy       2466 GATCCATTTGCCTCGGCCTCCCAAAGTGCTGGGAT-ACAGGCGTGAGCCACCATGCCCAA 2408
              ||||||   :::   ||||||||||||||||||   ||||||||||||||   |||:::
Db         37 AspProProSerLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHisGlyAlaArg 56

Qy       2407 CCTAGTTTCTCTACTTTAAAGCATGTATTTTAAAAATTATGGAACATCCTTAAATGGAAT 2348
              ||||||                                        :::|||:::

Db         57 ProSer-----------------------------------SerIleIle--------- 61

Qy       2347 ACTACATAGACATTTAGAAAAATTAGGTATGCCTGCACATGGAAATTTTCCCAAATAAAA 2288
                                                          |||:::     |||
Db         62 ------------------------------------------PheAsnSerGlnLys 66

Qy       2287 AAAGCTAGTGTATGCCCCAAAAATCTGGAGGAATATGTAACTGGTTATCTTTGGGTAGTA 2228
              :::   |||:::         ||||||::::::
Db         67 ArgMetSerIle---------AsnLeuAspLysGln------------------------ 75

Qy       2227 GGATTATAGATGCCTTCCACTTTCCATAATAGATATTTATGTAGAGTTTGAAGTTGCATT 2168
              |||:::     |||
Db         76 GlyValValHisPro---------------------------------------------- 80

Qy       2167 ACTTTTGGCTTGTTTACTTTTATAACCAAAATAAACAATAAGATTACAGTTAAAATGCAT 2108
                 :::   :::      |||:::   |||      :::
Db         81 ---TyrAsnIleLeuLeuPheValAsnLysLysGluSerAsnAsnAlaCysTyrAsnVal 99

Qy       2107 TTAAACTTTATTACATATAATTTACGTATCTCTGCTTACCTTCAGGAATCTTATTAGTTT 2048
              :::||| |||   :::|||:::                        ::::::   :::
Db        100 MetAsn-LeuGluAsnIleMetLeuLysLysLysProValThrLysAspGlnValGlyLe 119

Qy       2047 TGTTTCTCTCCGATCCTTTATTTTATTT------ATTTATTTTTTTGAGATAGAGTCTCG 1994
                 |||      |||:::||||||      :::   |||   |||      |||
Db        119 uPheValLeu------PheValLeuPheCysValValLeuPheIleGluAlaGlySerCy 137

Qy       1993 CTGTGTCGCCCAGGCTGGAGTGCAGTGGTGGGATCTCGGCTCACTGCAACCTCTGCCTCC 1934
                |||      |||||||||:::      ::: |||||||||:::|||   |||
Db        137 sSerAlaAlaHisProGlyValGlnPhe------Ile-AlaHisCysSerLeuHisLeuL 155

Qy       1933 CAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCATGCATC 1874
                 ||||||          |||||||||:::||||||   ||||||||||:::||||
Db        155 euAspSerSerProPro---ThrSerAlaSerGlnValAlaSerThrThrGlyValHisH 174

Qy       1873 ACTACACCCAGCTAATTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATATTGACCAGG 1814
              ||::::::|||||||||||      :::   :::   |||
Db        174 isHisSerGlnLeuIlePheAsnCysMetSerLysGlnGlyLeu---------------T 189

Qy       1813 CTGGTCTCGAACTTCTCACCTCAAGTGATCCACCCACCTTGGACTTCCAAAGTACTGTGG 1754
                |||      |||   |||:::           |||      ||||||      :
Db        189 hrValLeuProArgLeuValSerAsnSerSerGlnAlaLeuAlaSerGlnSerAlaArgI 209

Qy       1753 TTACAGGCTTGAGTCACTGTGCCCAACTGAGTTATTTTTAATTAATATGAACAAACATCC 1694
              ::||||||:::||||||      :::              |||      :::::::::
Db        209 leThrGlyValSerHisHisGlyArgProGluAspHisMetLeuPheAspSerHisLeu- 228

Qy       1693 TCTAATGCAGGTTTCTTGTCTTAGTTGCAAACATTAAAAAAAAACTGTGTGTGAGTGTGT 1634
                                                                  |||   |
Db        229 ------------------------------------------------GlySerAspC 232

Qy       1633 GTGTGTGTGTGTATTTCTCTGGGGAGAGGGTTCATAGCTTTCATCAGATTTGTAAATTTG 1574
              ||                        :::      |||   |||           |
Db        232 ys----------------------TrpLeuArgGlyPhePheGln-----------G 240

Qy       1573 GCCTT---CTAAAGTGGAGCAGGAGGATGTATAAGCAAATTAAG--------------C 1532
              |||||   |||:::|||                  ::::::|||             |
Db        240 lyLeuAsnLeuArgTrpLeuHisAsnSerValAsnArgValLysThrIleAspLeuCysP 260

Qy       1531 CATATTATTCAGTGACTAAAATAGTAACAATGAAAAAACAGTAACAATAAAGGTTTGAGG 1472
              ||      :::|||              |||              |||   |||:::
Db        260 roLeuAsnGluLeuThrValTrpTyrIleSerLysLeu-PheLysLysLysPheLeuLys 279

Qy       1471 CAATATGAAAAAGTCATAACAAGGTGATAAGTGAAAAAGCAGAATGCAAGATTGTATTTA 1412
                |||                                  |||   :::   :::
Db        280 LeuTyr-----------------------------------AsnLysLysArgPheAsn 287

Qy       1411 TTCTGCAAAGTTGTCAAGATTGTTTTAAAAAAA----------------GAAAAAGAA 1370
              :::   ||||||:::|||      :::   |||                ::::::|||
Db        288 TyrIleLysValIleLysThrLysIleAspLysAsnGlyArgIleAlaThrGlnGluGlu 307

Qy       1369 AGTTTATATCTA--------------------GGCTGTGCATTCATCAATTATTTATTG 1331
              |||   |||                        |||      ||||||
Db        308 SerTrpTyrLysValArgIleLeuIleCysAspSerCysLysIleIleAsn--------- 324
```

```
Qy       1330 AGCACCTACTATATGTCAGACACTGCTCTAGATCTGGGGATATAGCAATAAAGAAAATAG 1271
                                  :::   |||:::       |||   :::      |||::
Db        325 ---------------------SerLysLeuGluLys-AlaTyrGlyAspAsnHisAsnLy 337

Qy       1270 ACAGTAATCGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAG 1211
              :::: :::||||||      |||   |||       |||   |||   ||||||     |
Db        337 sArg-SerArgProAspAlaValAlaHisThrCysAsnProSerIleLeuGlyGlyArgG 357

Qy       1210 GCAGGCAGATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCC 1151
              ||      ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        357 lyGlyTrpIleThr---GlyGlnGluPheGluThrSerLeuGlyAsnMetValLysProA 376

Qy       1150 ATCTCTGCTAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGC 1091
                |||    |||      ||||||   |||   :::||||||   ||||||:::   |||
Db        376 rgLeu-TyrLysLysHisLysAsn---ProGlyValValValCysAlaCysSerProSer 394

Qy       1090 TACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGC 1031
                    ||||||   |||   ||||||
Db        395 CysLeuGlyGly---GlyThrArgIleAla---------------------------- 403

Qy       1030 TGAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAG 971

Db        403 ------------------------------------------------------------ 403

Qy        970 GCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATA 911

Db        403 ------------------------------------------------------------ 403

Qy        910 AGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGA 851
              :::   :::      |||||||||||||||||||||||||||||:::||||||:::|||
Db        404 ThrGlyGlnValArgTrpLeuThrProValIleProAlaLeuTrpGluAlaLysMetGly 423

Qy        850 GGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCT 791
              |||  :::      ||||||||||:::|||||||||   |||:::|||:::|||   |||
Db        424 Gly--SerArgGlyGlnGluPheGluThrSerLeuThrAsnMetValLysProHisLeuL 443

Qy        790 ACAAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTT 731
                  ||||||      ||||||:::|||||||||          |||:::|||||||||
Db        443 eu---LysIleGlnLysIleSerArgAlaTrpTrpArgValProValMetProAlaThrG 462

Qy        730 GGGAGGCTGAGGTCAGCAGATCCCTTGAGCT---------------------------- 700
                |||||||||      ||||||
Db        462 lnGluAlaGluAlaGlyGluLeuGluProGlyArgTrpArgLeuGlnAlaGluIleT 482

Qy        700 ------------------------------------------------------------ 700

Db        482 hrProLeuHisSerAsnLeuGlyAsnLysAlaArgLeuArgLeuLysLysLysLysL 502

Qy        699 -----------------CAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCACCACTGTGC 657
                              :::|||   :::   ||||||:::      |||||||||   |
Db        502 ysLysArgIleAlaIleArgGluGluLysValAlaValSerArgAspCysThrThrAlaL 522

Qy        656 TCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAACAACAAAAAAAAAACCCAAA- 598
              ||      |||   ||||||:::|||||||||   |||||||||   :::|||   |||
Db        522 euGlnSerGly---GlnSerGluThrLeuSerGlnAsnAsnAsnAsnArgLysHisLysL 541

Qy        597 -----------------CTGTCACCACTTACAAGC---------------TGTGTGATCT 570
                               |||:::   :::||||||                 ::::::
Db        541 ysGlnIleAsnLysGluLeuAsnMetValThrSerGluTyrGlySerGluLeuLeuValS 561

Qy        569 TGGGCAAATAGCTCAGCCTTCTGAGCCTCAGTTTTCTT-----------------ATCT 528
                                 |||   ::::::|||:::                  :::|
Db        561 erTyrAlaProPheAspTyrLeuCysIleAsnPheIleAsnHisIleAsnSerHisValC 581

Qy        527 GTTAAGTGAGGATAATTGTACCTATCCCATAGTGTTGTTCTGAGGATTAAAGGAGACACT 468
              ||   :::   ::::::   |||
Db        581 ys---IleTyrMetLeuPhePro-----------------------------------T 588

Qy        467 GCCCAGAACCAAGAAAATTAACACTAAATCATGGCTGT-----------TGTTACTATT 420
                ||:::   :::|||            :::||||||            |||   :::|
Db        588 rpProLysArgLysLysSerLysSerLysTyrGlyCysValIleGlnLeuCysLysPheT 608

Qy        419 ATTTCCCCCCGGCCTAC-----------TGAGAACCAACTAGAATTCATAGGTCAAGGA 372
                ||      |||   :::               :::   ||||||:::   |||:::
Db        608 yrCysIleProLeuPhePhePheSerAsnSerLysLeuThrArgLeuSerArgThrThr- 627

Qy        371 AGGGATAAACCTACTGCTACAGGTGATTGCCGAGCCAGAGCACGAGTATGAATAGGTGTG 312
                      |||      ::::::   :::          ||||||
Db        628 --------ThrLeuAlaHisLysLysIleLeuGluHisSerThr-----------PheA 641

Qy        311 GGTGTACGGKKTGTCCAGCAGAAATT---------TACAGCTGTCCAGCTTCTTGGCCTG 261
                ||||||         :::|||:::        :::         ::::::
Db        641 laCysThrValHisLeuSerGluValThrArgAspPheLeuAlaAlaSerAlaAlaSerA 661

Qy        260 GTCRTAGCAGTTGTCATGTG-----------------------------------TCTG 237
                  :::|||      |||
Db        661 laGlnAsnSerLeuLeuValLeuThrGluValGlyValLysGlyLeuAlaGluSerLysS 681

Qy        236 GCAGCACCTGGAAAGTGGGA---GGGACAGCTGAGATAGGAGTAAGTGCAGAGCAATAGG 180
                |||   |||   :::|||   |||   |||   :::   |||          |||   |
Db        681 erSerArgTrpAspIleGlyArgGlyGlnAlaTrpValLysValHisGlnIleGlnLysV 701

Qy        179 TCAGCCCTCACCTGCCCACTCTCAGGAACAGGTGGGGATGACTTCGCCAAGATGCTTCAG 120
              ||:::      |||                                           |
Db        701 alAlaArgGlnLeu-----------------------------------ThrProGluG 709

Qy        119 GAGAAATGATCCTATGGCTTGAA 97
                ||:::   :::   |||||||||
Db        709 lyLysLysValArgTrpLeuGlu 716
```

RESULT 5
AAU31513
ID AAU31513 standard; protein; 215 AA.
XX
AC AAU31513;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2004.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 452; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered

```
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 215 AA;

Alignment Scores:
Pred. No.:              7.16e-42      Length:        215
Score:                  438.50        Matches:       116
Percent Similarity:     59.9%         Conservative:  14
Best Local Similarity:  53.5%         Mismatches:    60
Query Match:            9.9%          Indels:        28
DB:                     4             Gaps:          3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAU31513 (1-215)

Qy       639 TCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCT 698
Db         4 SerValThrGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 23

Qy       699 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 758
Db        24 ArgPheArg***PheSerCysLeuSerLeuLeuSerSerTrpAspTyrArgArgProPro 43

Qy       759 CCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAG 818
Db        44 ProArgProAlaAsn-PheCysIlePheSerArgAspMetAlaPheThrThrLeuAlaAr 63

Qy       819 GCTTGTCTTAAACTCCTGGCTTCAAGC-AATCCTCCTACCTTGGCATCCCAAAGTGCTGG 877
Db        63 gLeuValSerAsnSer***ProGlnCysAspProProThrSerAlaSerGlnSerAlaGl 83

Qy       878 GATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAAT 937
Db        83 uIleThrGlyValSerHisArgAlaTrpProValLeuSerSerProGlnProPhe----- 101

Qy       938 AGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGAC 997
Db       102 ----------------------------------------------PhePhePheAspMe 106

Qy       998 AGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC--------------TCACTGCAAC 1042
Db       106 tGluSerHisAlaIleThrGlnAlaGlyVal-GlnTrpArgHisLeuGlySerLeuGlnP 126

Qy      1043 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 1102
Db       126 roProProProMetPheLys***SerSerCysLeuSerLeuLeuSerSerTrpAspTyrA 146

Qy      1103 AGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTT---AGCAGAGATGGGGTTTTA 1159
Db       146 rgArgProProProArgProAlaAsnPhePhe-ValPheLeu***ArgAspGlyValSer 165

Qy      1160 CCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTC 1219
Db       166 ProCys***ProGlyTrpSerArgSerProAspLeuValIleHis-ProProTrpSerPr 185

Qy      1220 CCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
Db       185 oLysSerAlaGlyIleThrGlyLeuSerHisCysAlaGlnPro 199

RESULT 6
ADC87151
ID   ADC87151 standard; protein; 217 AA.
XX
AC   ADC87151;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR protein SEQ ID NO:1604.
XX
KW   human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   N-PSDB; ADC87150.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 2; SEQ ID NO 1604; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The protein
CC   sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ   Sequence 217 AA;

Alignment Scores:
Pred. No.:              1.44e-40      Length:        217
Score:                  427.50        Matches:       115
Percent Similarity:     52.3%         Conservative:  11
Best Local Similarity:  47.7%         Mismatches:    64
Query Match:            9.7%          Indels:        51
DB:                     7             Gaps:          5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADC87151 (1-217)

Qy       499 TGGGAT-----------------AGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
Db        17 TrpAspGlnSerServalGlnPheArgValAsnPheLeuLeuLeuValAsp--------- 33

Qy       541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
Db        34 --------------------PheThrLysMetLysAsnPheCysSerSerGluAsnIle 46

Qy       601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
Db        47 LeuPhePhePheLeu---------ArgArgSerLeuThrLeuLeuProArgLeuGluCys 63

Qy       661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
Db        64 SerGlyThrIleSerAlaHisCysAsnLeuHisLeuProGlySerGlyAspSerHisThr 83

Qy       721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
Db        84 SerAlaSerGlnValAlaGlyIleThrGlyAlaHisArgHisAlaTrpLeuIlePheVal 103

Qy       781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
Db       104 --PheLeuLeuGluThrGlyPheCysHisValGlyGlnAlaGlyLeuGluLeuLeuThrS 123
```

```
Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||   :::||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        123 erGlyAspProProAlaLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHis---- 141

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960

Db        141 ------------------------------------------------------------ 141

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
                          |||       |||||||||||||||||||||||   :::|||
Db        142 -----ProLysProPhePhePhePhePheGluThrGluSerHisThrValThrTrpA 160

Qy       1021 CCTCCGTTCAGC--------------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGA 1065
                  |||                 ||||||||||||   ||| |||||||||||||
Db        160 laGlyVal-GlnTrpCysAspLeuGlySerLeuGlnProProPro-ProGlySerSerAs 179

Qy       1066 TTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCT 1125
              ||||||||||   ||||||   ||||||||||||||      ||||||||||||   ||
Db        179 pSerProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaHisHisHisAlaGlnLe 199

Qy       1126 AATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCT 1180
              |||||||      |||   ||||||||||||      |||   ||||||||||||
Db        199 uIlePheLeuPhePheArgArgAspGlyValSerLeuCysTrpProGlyTrpSer 217

RESULT 7
AAE35265
ID   AAE35265 standard; protein; 7067 AA.
XX
AC   AAE35265;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Human P450RAI-3 protein #2.
XX
KW   Human; retinoic acid; vitamin A; cytochrome P450; actinic keratosis;
KW   P450RAI; cancer; psoriasis; acne; ichthyosis; gene therapy; vaccine;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200295034-A2.
XX
PD   28-NOV-2002.
XX
PF   23-MAY-2002; 2002WO-CA000758.
XX
PR   23-MAY-2001; 2001US-0292531P.
XX
PA   (CYTO-) CYTOCHROMA INC.
XX
PI   Wisniewski J,  Petkovich PM,  Ramshaw H;
XX
DR   WPI; 2003-112152/10.
DR   N-PSDB; AAD53861.
XX
PT   New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
PT   diagnosing, preventing, or treating disorders with aberrant expression or
PT   activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
PT   psoriasis.
XX
PS   Example 1; Fig 1; 231pp; English.
XX
CC   The invention relates to retinoic acid metabolising cytochrome P450,
CC   P450RAI polypeptides and polynucleotides. Methods and compositions of the
CC   invention are useful for diagnosing, preventing, ameliorating and/or
CC   treating disorders associated with the aberrant expression or activity of
CC   the P450RAI such as diseases related to vitamin A and retinoic acid
CC   metabolism, e.g. cancer, actinic keratosis, psoriasis, acne and
CC   ichthyosis. The invention is useful in gene therapy and as vaccine. The
CC   present sequence is human P450RAI protein
XX
SQ   Sequence 7067 AA;
```

```
Alignment Scores:
Pred. No.:              4.71e-39        Length:         7067
Score:                  420.50          Matches:        172
Percent Similarity:     40.0%           Conservative:   51
Best Local Similarity:  30.9%           Mismatches:     148
Query Match:            9.5%            Indels:         188
DB:                     6               Gaps:           20

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAE35265 (1-7067)

Qy        737 CTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGAC 796
              |||||||||:::          |||   ||||||              ||||||   ||||||
Db          2 LeuGlyLeuGlnValProAlaThrMetProGlyPhePheIleTyrPhePheSerArgAsp 21

Qy        797 TGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTAC 856
                 |||   |||   :::||||||   |||      |||   :::      |||      |
Db         22 ArgValSerLeu---TyrProGlyTrpSer--ArgSerHisValIleMetIleHisProP 40

Qy        857 CTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTC 916
              |||||   ||||||||||||||||||||||      ||||||:::   |||          |
Db         40 roTrpProProLysValLeuGlyLeuGlnAla---AlaThrValProGlyProAlaAlaL 59

Qy        917 TTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTT 976
              ||::::::   :::   :::            ||||||:::     :::      ::::
Db         59 euPheLeuIleAlaLysLysAspLysCysProSerThrAspAspTrpIleAsnLysValT 79

Qy        977 TCTTCCTT---------------------------------------------------- 984
              ::::::::
Db         79 rpTyrIleHisThrIleTyrTyrTyrLeuSerIleLysArgAsnGluValIleMetHisA 99

Qy        984 ------------------------------------------------------------ 984

Db         99 laThrThrThrLeuLysIleLeuCysLeuSerGluArgSerGlnLeuGlnLysThrArgV 119

Qy        985 ----------------------------------TCTTTTTTGAGACAGAGTCTCACTT 1009
                                                 |||   ||||||   |||
Db        119 alPheLeuPheCysLeuPheCysPheValLeuPheCysPheLysArgGlnGlyLeuAlaL 139

Qy       1010 CATCACCTCAACCTCCGTTCAGCTCACTGCA-ACCTCTGCCTCCCGG------GTTCAAG 1062
                 |||         :::|||||||||   |||:::||||||          |||
Db        139 euLeuProIleLeuGluTyrSerSerLeuLeuThrAlaAlaSerThrPheTrpThrGlnA 159

Qy       1063 YGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCAT----G 1118
                 |||   |||:::||||||:::   |||   ||||||   |||||||||      :
Db        159 laSerLeuProProGlnProProLys---LeuAlaLeuGlnValCysThrThrThrAlaS 178

Qy       1119 CTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGT 1178
              :::::||||||   |||         |||:::   |||   ||||||||||||||||||
Db        178 erPheLeuIleValCys---------ArgAsnArgValSerProCysCysProGlyTrpS 195

Qy       1179 CTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATA 1238
              ||||||||||   ||||||       |||      |||           |||:::
Db        195 erGlnThrProSerLeuLys-ProTrpProLeuLysVal---------LeuGlyLeuGln 211

Qy       1239 GGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGA 1298
              |||   :::                  |||   |||                 ||||||
Db        212 GlyAlaThrMetValGlyGlnLysThrThrCysTyrLeu------------IleProIle 227

Qy       1299 TCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCT 1358
                    :::::::::
Db        228 TrpGlyValIleAlaAsn---------------------------------------- 233

Qy       1359 AGATATAAACTTTCTTTTTCTTTTTTTAAA--------------ACAATCTTGACAACT 1403
                 |||   :::||||||         |||              |||||||||
Db        234 GlyTyrGlyValSerPheArgGlySerLysIleAspAspGlyCysThrIleLeu------ 251

Qy       1404 TTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTT 1463
              :::::::::|||      :::                  |||           |||
Db        252 IleGluSerLysProLeuIleCysAlaLeuMetSerLeuTyrGly---------ThrSer 268

Qy       1464 TCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGA 1523
                 |||||||||                        :::   ::: :::|||
Db        269 GlnSerCysLeuLys--------------------LysAsnPheSerTyr-IleThrAr 281
```

```
Qy       1524 ATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACA 1583
                 |||||||||         ||||||   |||     :::         :::
Db        281 gAsnArgGlyLeuIle---------IleLeuLysLeuLysGlnLysThrLysMetValGl 298

Qy       1584 AATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCAC 1643
                :::                         |||:::|||                  :::
Db        298 yLeu-----------------------ProLysLysArgAla---------GlyIleLy 307

Qy       1644 ACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCAT-------------- 1689
                    |||      :::   |||   ||||||:::|||   :::
Db        307 sGlyLysPheSerTyrValIleAlaValLysThrGlnAsnLysAsnLysProIleArgAs 327

Qy       1690 -TAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCC 1748
                  :::                                              |||  |||   ||
Db        327 pSerGluIleIleThrArgThrLysArgGlyValLysGlyGlnThrGlnTrpLeuThrPr 347

Qy       1749 TGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAG 1808
                 ||||           |||:::::::::|||||||||||||||||||   ||||||   ||
Db        347 oValIleProAlaPheTrpGluAlaGluValGlyGlySerLeuGluAlaArgSerLeuAr 367

Qy       1809 ACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGT 1868
                 ||||||||||:::   |||   ||||||:::|||||||||||||||||||||||||
Db        367 gProAlaTrpAlaThrTrp---AsnProValSerThrLysAsnThrLysIleSerArgAl 386

Qy       1869 GTAGTGATGCATGCCTGTAGTCCCAGCTACT---------------------------- 1899
                        :::|||||||||||||||
Db        386 aTrpTrpTyrValProValValProAlaAlaTrpGluAlaGluAlaGlnGluLeuLeuGl 406

Qy       1899 ------------------------------------------------------------ 1899

Db        406 uProGluAlaArgCysGlyGlySerArgLeuSerGlnHisPheGlyArgProArgTrpVa 426

Qy       1899 ------------------------------------------------------------ 1899

Db        426 lAspHisGluValArgSerLeuArgProAlaProThrTrpAsnProIleSerAsnLysTy 446

Qy       1900 ----------------------------------------------CGGGAGGCTGAGGC 1913
                                                          ||| ||||||||||
Db        446 rLysLysLeuAlaGlyHisGlyGlyValCysLeuCysGlnLeuLeuArg-ArgLeuArgG 466

Qy       1914 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
                 ||||||||||||||||||||||||   ||||||||||||||||||||   |||||||
Db        466 lnGluAsnCysLeuAsnLeuGlyGlyGlyGlyCysSerGluProArgSerArgHisCysT 486

Qy       1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAA 2014
                 |||||   |||||||||:::|||||||||:::|||||||||
Db        486 hrProThrTrpAlaThrLysArgAspSerValSerLysLys 499
```

RESULT 8
ADA84009
ID ADA84009 standard; protein; 673 AA.
XX
AC ADA84009;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human POM77 protein.
XX
KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;
KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
KW vaccine.
XX
OS Homo sapiens.
XX
PN WO2002103028-A2.
XX
PD 27-DEC-2002.
XX
PF 30-MAY-2002; 2002WO-IB004189.
XX
PR 30-MAY-2001; 2001US-0293999P.
PR 22-OCT-2001; 2001US-0330457P.
PR 19-FEB-2002; 2002US-0357144P.
XX
PA (BIOM-) BIOMEDICAL CENT.
XX
PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;
XX
DR WPI; 2003-175241/17.
DR N-PSDB; ADA84008.
XX
PT Determining if a nucleic acid is a marker for a phenotype/cell type of
PT interest, by global comparison of expressed sequence tags known to be
PT expressed in the phenotype/cell type with all ESTs expressed in normal
PT tissue.
XX
PS Claim 29; Page 392-395; 516pp; English.
XX
CC The invention relates to a novel method for determining if a nucleic acid
CC is a marker for a predetermined phenotype/cell type of interest from a
CC biological species. The method comprises performing a global comparison
CC of a group of expressed sequence tags (ESTs) known to be expressed in the
CC phenotype/cell type of interest with all ESTs expressed in normal tissue
CC in order to identify ESTs that are preferentially expressed in the
CC phenotype/cell of interest. A method of the invention is useful for
CC determining whether a nucleic acid is a marker for a predetermined
CC phenotype or cell type of interest from a biological species, preferably
CC Arabidopsis or human. The cell type of interest is an abnormal cell such
CC as a tumour cell, and the predetermined phenotype is a stress-induced
CC phenotype such as hyperosmotic stress or high salt conditions. A method
CC of the invention is also useful for determining the progression of colon
CC cancer in a human, for detecting a tumour cell, and for regulating or
CC preventing the growth of a tumour cell. An antibody of the invention is
CC useful for detecting the absence or presence of peptides encoded by
CC tumour-associated markers. A polypeptide of the invention is useful as an
CC immunogen for vaccinating an animal. The present sequence represents a
CC tumour-associated antigen of the invention.
XX
SQ Sequence 673 AA;

Alignment Scores:

Pred. No.:	5.53e-39	Length:	673
Score:	416.00	Matches:	131
Percent Similarity:	41.8%	Conservative:	17
Best Local Similarity:	37.0%	Mismatches:	85
Query Match:	9.4%	Indels:	122
DB:	6	Gaps:	9

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADA84009 (1-673)

```
Qy        532 AAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGG 591
                  ||||||         ||||||               |||         |||   |||
Db        328 LysLysLysAsnProGlnLys--------------Gln*********PheCysPheTrp 342

Qy        592 TGACAGTTTGGGTTTTTTTTTGTTGTTGTTTA-------GAGACAGGGTCTTGCTCTGTC 644
                 |||   |||||||||       |||          |||   |||   ||||||
Db        343 GlyPhePhePhePhePheCysPheValPheValPheGluIleGlyProCysSerAsp 362

Qy        645 ACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTC 704
                 |||   |||       ||||||        :::   ||||||||||||||||||||   ||
Db        363 ThrProAlaGlyValGlnTrp-GlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySe 382

Qy        705 AAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGC 764
                 |   |||           |||   |||:::|||||||||   |||   :::       |||::
Db        382 rSerAspSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSe 402

Qy        765 CTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGT 824
                 :|||||||||         |||   :::||||||       ::::::|||   ||||||
Db        402 rTrpLeuIleTyrVal--PheLeuIleGluThrGlnPheHisHisValAspGlnAlaGly 421

Qy        825 CTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACA 884
                 ||||||||||||   ||||||:::   |||::: |||   ||||||||||||||   ||
Db        422 LeuLysLeuLeuThrSerSerAspLeuProSer-TrpSerProLysValLeuGly***Gl 441

Qy        885 GGGGTGAGCCACC-----------------------ATGTGCGGCTACTTATTTCTTTA 920
```

```
Db        441 nAla***AlaThrThrProSer*********PheValPheGlyGlyPhePhePhePhePh 461

Qy        921 CATTCCA---------------------------------------------------- 927
              :
Db        461 ePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArgLeuGluCysSe 481

Qy        928 ----TCTTTCCAATAGAATGTAAGATCCACA---------GAACAGGGATTACTGCCTAT 974
                  |||                 ||||||        :::      |||||||||::
Db        481 rGlyLysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLe 501

Qy        975 T-----------------------------------------------TTCTTCCTTTC 986
              :                                                |||
Db        501 uSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyrPh 521

Qy        987 TTTTTTGAGACAGAGTCTCACT----------------------TCATCACCTCAA-- 1020
                  |||   |||   |||                        |||   ||||||
Db        521 e******ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVa 541

Qy       1021 ---------------------------------------------CCTCCG------- 1026
                                                           ||||||
Db        541 lIleCysProProGlyLeuProLysCysTrpAspAspArgArgGluProProHisProAl 561

Qy       1026 ------------------------------------------------------------ 1026

Db        561 a*********LeuPheLeuGlyValPhePhePhePhePheLeuLeuCysPheCysPhe** 581

Qy       1027 ---------------------------------------------TTCAGCTCACTGCA 1040
                                                           |||   ||||||||
Db        581 *AspArgProLeuLeu***HisProGlyTrpSerAlaValAlaSerPheGlySerLeuGl 601

Qy       1041 ACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTA 1100
              ||||   ||||||||||||||   ||||||||||||   |||      ||||||:::||
Db        601 nProGlnProProGlyPheLysArgPheSerCysLeuCysLeuProCysSerTrpAspTy 621

Qy       1101 CAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTAC 1160
              |:::      ||||||   |||||||||    |||||||||:::||||||   |||   |
Db        621 rArgHisGlyProProPheLeuAlaAsn-LeuCysIlePheAsnArgAspThrValSerP 641

Qy       1161 CATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCC 1220
              |||||      ||||||||||||||||||||||||   ||||||   ||| ||||||
Db        641 roCys***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSer 660

Qy       1221 CAAAGTGCTGGAATTATAGGCGTGAGTCAC 1250
              ||||||||||||:::   ||||||||||||
Db        661 GlnSerAlaGlyMetThrGlyValSerHis 670
```

```
RESULT 9
AAO06014
ID   AAO06014 standard; protein; 185 AA.
XX
AC   AAO06014;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 19906.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI85945.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 20; SEQ ID NO 19906; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 185 AA;
```

Alignment Scores:

Pred. No.:	6.98e-39	Length:	185
Score:	413.00	Matches:	98
Percent Similarity:	54.7%	Conservative:	12
Best Local Similarity:	48.8%	Mismatches:	49
Query Match:	9.3%	Indels:	42
DB:	4	Gaps:	3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAO06014 (1-185)

```
Qy        642 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              :::   ||||||      ||||||||||||      ||||||||||||   ||||||
Db         11 IleAlaGlnAlaGlyValGlnTrpCysAsnLeuSerSerLeuGlnProLeuProProArg 30

Qy        702 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 761
                 |||      :::   |||||||||||||||||||||||||||          ||||||
Db         31 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisValProPro 50

Qy        762 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 821
              ||||||||||||          |||   ||||||   |||      :::||||||||||||
Db         51 ArgLeuAlaAsnPhePheValPheLeuValGluMetGlyPheLeuHisValGlyGlnAla 70

Qy        822 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACC---------------------TTG 860
              |||:::||||||||||||   :::||||||:::                       |||
Db         71 SerLeuGluLeuLeuAlaSerGlyAspProProSerGlyAspProAspLeuArg***Leu 90

Qy        861 GCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTA 920
              |||   ||||||   |||||||||||||:::||||||     |||   :::||||||::
Db         91 AlaPheGlnSerLeuGlyIleThrGlyValThrHisHis-AlaGlyGlnIlePheLeuPh 110

Qy        921 CATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT 980
              :
Db        110 eCysLeuGlu---------------------------------------------- 113

Qy        981 CCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCG------------- 1026
                                  |||   ||||||
Db        114 ------------------------ThrGluSerProSerValAlaGlnValGlyValGl 125

Qy       1027 ----------TTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 1076
                       ||||||||||||||||||||||||   :::||||||||||||||
Db        125 nTrpCysAsnLeuGlySerLeuGlnProLeuProProGlySerGlu***PheSerCysLe 145

Qy       1077 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTA 1136
              |   |||   |||||||||:::|||:::   |||          ||||||  |||   |
Db        145 uGlyLeuProSerSerTrpAspTyrArgHisAlaLeuThrHisProAlaAsn-PheSerI 165
```

```
Qy       1137 TTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195
              ||||||||||||||||||||   |||     |||||||||||||||   |||||||||
Db        165 lePheSerArgAspGlyValSerProSerTrpProGlyTrpSerGlnMetProAspLeu 184

RESULT 10
ADK36978
ID   ADK36978 standard; protein; 232 AA.
XX
AC   ADK36978;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID9060.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .232
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 9060; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 232 AA;

Alignment Scores:
Pred. No.:              6.85e-38      Length:         232
Score:                  405.00        Matches:        108
Percent Similarity:     56.1%         Conservative:   11
Best Local Similarity:  50.9%         Mismatches:     52

Query Match:            9.2%          Indels:         42
DB:                     5             Gaps:           5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADK36978 (1-232)

Qy        600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
              |||   ||||||           :::|||   |||   ||||||||||||||
Db         52 TrpProPhePhe------PhePheGlnThrGluSerArgSerValThrGlnAlaGlyVal 69

Qy        660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
              |||   ||||||      ||||||||||||   ||||||   ||||||      :::
Db         70 Gln***CysAsnLeuSerSerLeuGlnProLeuProProGlyLeuLysGlnPheSerArg 89

Qy        720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
              |||||||||||||||||||||||||||          |||||||||   |||:::
Db         90 LeuSerLeuProSerSerTrpAspTyrArgHisMetProProArgProAlaSer---Phe 108

Qy        780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG-GC 838
                |||   |||||||| |||   |||   ||||||||||||||   ||||||
Db        109 ValPheLeuValGluThr-GlyPheThrMetLeuAlaArgLeuValTrpAsnSer***Pr 128

Qy        839 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC----- 893
                |||   :::||||||      |||||||||:::|||||||||||||||   |||
Db        128 oSerCysAspProProIleSerAlaSerGlnAsnAlaGlyIleThrGly***SerProSe 148

Qy        894 ----------CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATG 943
                       ||||||   ||||||
Db        148 rLeuAlaProHisHisProArgLeu---------------------------------- 156

Qy        944 TAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTC 1003
                                                  ||||||||||||||:::||
Db        157 ----------------------------------------PhePhePheGluThrLysSe 163

Qy       1004 TCACTTCATCACCTCAACCTCCGTTCAGC--------------TCACTGCAACCTCTGC 1048
              |      :::|||          |||                  |||   ||||||||||
Db        163 rGlySerValThrGlnAlaGlyVal-GlnTrpHisAsnProThrSerGlnGlnProLeuP 183

Qy       1049 CTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTG 1108
              |||||||||||||      |||   |||||||||   |||||||||:::   :::   |
Db        183 roProGlyPheLysGlnLeuSerArgLeuSerLeuSerSerSerTrpAsp***ArgTyrA 203

Qy       1109 CACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGC 1168
              |||||   |||||||||||| :::||||||||||||||||||         ||||||
Db        203 laProSerCysLeuAlaAsn-TyrCysIlePheSerArgAspArgAlaSerProCysTrp 222

Qy       1169 CCAGGCTGGTCTCAAACTCCTGACCTCAAG 1198
              |||||||||   ||||||||||:::
Db        223 SerGlyTrpSer***ThrProAspLeuArg 232

RESULT 11
ADK36936
ID   ADK36936 standard; protein; 302 AA.
XX
AC   ADK36936;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID9018.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .302
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
```

```
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 9018; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 302 AA;

Alignment Scores:
Pred. No.:              3.02e-37        Length:         302
Score:                  400.00          Matches:        110
Percent Similarity:     53.1%           Conservative:   20
Best Local Similarity:  44.9%           Mismatches:     72
Query Match:            9.0%            Indels:         43
DB:                     5               Gaps:           5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ADK36936 (1-302)

Qy        597 GTTTGGGTTTTTTTTTGTTGTTGT----------TTAGAGACAGGGTCTTGCTCT 641
              |||      ||||||      |||               |||   ||||||     :::
Db         81 ValCysThrPhePheIleTyrCysProPheLeuPhePheLeuArgThrGlyPheSerAla 100

Qy        642 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              ::::::||||||     ||||||   :::     ||||||||||||   ||||||
Db        101 LeuSerGlnAlaValValGlnTrpArgAspLeuSerSerLeuGlnProProProProGly 120

Qy        702 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 761
              ||||||      :::   ||||||||||||||||||||||||:::         ||||||
Db        121 LeuLysArgPheSerCysLeuSerLeuProSerSerTrpAspHisArgHisValProPro 140

Qy        762 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 821
                  ||||||      |||   :::   |||   |||   |||     ||||||::
Db        141 HisProAlaAsn-PheCysIleLeuMet***ArgTrpGlyPheThrMetPheAlaArgMe 160

Qy        822 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 881
              :      :::               :::|||  |||   |||||||||:::     |||
Db        160 tThrCysAsp--------------LeuLeu--ThrSerAlaSerGlnThrValArgIle 174

Qy        882 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 941
              ||||||||||||||||   ||||||
Db        175 ThrGlyValSerHisHisThrArgLeu-------------------------------- 183

Qy        942 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTT----------CTTCCTTTC 986
                                  |||      ||||||           :::   |||
Db        184 --------------------IleTyr***TyrPhe***Ala***SerAspIlePhePhe 196

Qy        987 TTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC---------- 1032
              |||||||||   ||||||     :::          |||
Db        197 PhePheGluLysGluSerCysCysValGlyGlnAlaGlyVal-GlnTrpHisAspLeuGl 216

Qy       1033 -TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAG 1091
               ||||||||||||   |||  |||  |||  ||||||||||:::|||  |||||
Db        216 ySerLeuGlnProProProSerGlyLeuLysGlnPheSerCysLeuAsnLeuProSerSe 236

Qy       1092 CTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATG 1151
              ||||::::::||||||   ||||||         ::: ||||||||||||||||||||
Db        236 rTrpAspHisLysArgThrProProHisProGlySer-PheCysIlePheSerArgAspA 256

Qy       1152 GGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCC 1211
                |||   |||      |||||||||      ||||||||||||||   ||||||   |
Db        256 rgValSerProTrpGlyProGlyTrpPhe***ThrProAspLeuLysArgSerAlaLeuL 276

Qy       1212 TCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTC 1271
              ||   ||||||||||||   :::||||||||||||           |||:::     :
Db        276 euGlyLeuProLysCysCysAspTyrArgArgGluProProArgProAlaGluValThrI 296

Qy       1272 TATTT 1276
              ::|||
Db        296 lePhe 297

RESULT 12
ABG68738
ID   ABG68738 standard; protein; 361 AA.
XX
AC   ABG68738;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human prostate specific protein DEX0293_107.
XX
KW   Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
KW   prostate specific protein; cytostatic; non-cancerous prostate disease;
KW   gene therapy; cancer; immunostimulant; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200255735-A2.
XX
PD   18-JUL-2002.
XX
PF   27-NOV-2001; 2001WO-US044363.
XX
PR   27-NOV-2000; 2000US-0253176P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Recipon H,  Cafferkey R,  Ali S,  Sun Y;
PI   Liu C,  Chen S;
XX
DR   WPI; 2002-557831/59.
DR   N-PSDB; ABK97631.
XX
PT   New prostate specific genes, useful for treating or diagnosing cancer, or
PT   useful as vaccines for treating cancer, particularly prostate cancer, in
PT   a patient.
XX
PS   Claim 11; Page 207-208; 212pp; English.
XX
CC   The invention relates to a new isolated prostate-specific nucleic acid
CC   (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
CC   ABK97642 which encode prostate specific proteins appearing as ABG68701-
CC   ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
CC   homology with a PSNA. Also included are a method of determining the
CC   presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
CC   comprising the vector, producing the polypeptide encoded by the PSNA, a
```

```
CC   method of determining the presence of a PSP in a sample, diagnosing and
CC   monitoring the presence and metastases of prostate cancer in a patient, a
CC   kit for detecting a risk of cancer or presence of cancer in a patient
CC   (the kit comprising a means for determining the presence of the PSNA or
CC   PSP in a sample of a patient) and a vaccine comprising the polypeptide or
CC   the nucleic acid encoding the polypeptide. The PSNA, PSP and anti-PSP
CC   antibody are useful for diagnosing and treating cancer in a patient (e.g.
CC   by gene therapy). The nucleic acid molecule and polypeptide are also
CC   useful as vaccines for treating cancer, particularly prostate cancer and
CC   non-cancerous prostate diseases. The present sequence is a PSP of the
CC   invention
XX
SQ   Sequence 361 AA;

Alignment Scores:
Pred. No.:              3.27e-37      Length:         361
Score:                  400.00        Matches:        105
Percent Similarity:     55.8%         Conservative:   11
Best Local Similarity:  50.5%         Mismatches:     55
Query Match:            9.0%          Indels:         38
DB:                     5             Gaps:           5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ABG68738 (1-361)

Qy        639 TCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCT 698
              ||||||  ||||||     ||||||||||||     ||||||||||||   ||||||
Db        172 SerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuProPro 191

Qy        699 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 758
                  :::     :::   |||||||||   ||||||||||||:::     ||||||
Db        192 ArgPheArgArgPheSerCysLeuSerLeuLeuSerSerTrpAspHisArgHisAlaPro 211

Qy        759 CCACGCCTGGCTAATTAAAAA------AATTTTTTTGTAGAGACTGGGTCTTACTACGTT 812
              |||   |||||||||               |||   |||:::          |||   ||
Db        212 ProCysLeuAlaAsnPheLeuPhePheLysPheLeuValAspGln-SerPheThrMetLe 231

Qy        813 GGCCAGGCTTGTCTTAAACTCCTG-GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAG 871
              ||||||||||||||||||||||      |||   :::   |||      ||||||||||
Db        231 uAlaArgLeuValLeuAsnSerAlaProSerGlyAspLeuProAlaProAlaSerGlnSe 251

Qy        872 TGCTGGGATTACAGGGGTGAGCCAC---------------------------------- 896
              ||||||||||||||||||   |||
Db        251 rAlaGlyIleThrGlyValArgHisCysThrTrpProLysSerPheLysPheAlaAspSe 271

Qy        897 -CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAG 955
               |||:::   |||      :::   |||      |||
Db        271 rHisIleGlyLeu------AlaPheHisPheAlaPhe---------------------- 281

Qy        956 AACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCAC 1015
                          :::|||      |||||||||          ||||||   |||
Db        282 ---------------PhePhePhePhePhePheAlaValAlaSerHisProIleAl 296

Qy       1016 CTCAACCTCCGTTCAGC--------------TCACTGCAACCTCTGCCTCCCGGGTTCA 1060
                        |||                  ||||||||||||   ||||||||||||
Db        296 aGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnProProProProGlyPheL 316

Qy       1061 AGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCT 1120
              ||   |||   ||||||||||||      ||||||:::|||:::||||||||||||
Db        316 ysGlnPheLeuCysLeuSerLeuProGlySerTrpAspTyrArgArgAlaProProArgG 336

Qy       1121 TGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCT 1180
                ||||||  ||||||||||||||||||||||||||||   ||||||      |||||||||
Db        336 lnAlaAsn-PheCysIlePheSerArgAspGlyValSerProCysTrpThrGlyTrpSer 355

Qy       1181 CAAACTCCTGACCTCAAG 1198
              ||||||||||||||:::
Db        356 GlnThrProAspLeuArg 361

RESULT 13
ABJ01057
ID   ABJ01057 standard; protein; 449 AA.
XX
AC   ABJ01057;
XX
DT   28-NOV-2002  (first entry)
XX
DE   Ovary cell-specific amino acid sequence 3.
XX
KW   Ovary cell; neoplastic ovary cell; ovary specific nucleic acid;
KW   ovary specific protein; ovarian cancer; breast cancer; vaccine;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200238606-A2.
XX
PD   16-MAY-2002.
XX
PF   07-NOV-2001; 2001WO-US046459.
XX
PR   08-NOV-2000; 2000US-0246640P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y,  Recipon H,  Salceda S,  Liu C;
XX
DR   WPI; 2002-519297/55.
XX
PT   Polypeptide and polynucleotides present in normal and neoplastic ovary
PT   cells, useful for identifying, monitoring, staging, diagnosing,
PT   preventing and treating ovarian cancer, and non-cancerous disease states
PT   in the ovary.
XX
PS   Claim 11; Page 207-209; 247pp; English.
XX
CC   The invention comprises amino acid and DNA sequences which are present in
CC   normal and neoplastic ovary cells. The DNA and protein sequences of the
CC   invention are useful for determining the presence of an ovary specific
CC   nucleic acid or an ovary specific protein in a sample. The DNA and
CC   protein sequences of the invention are useful for diagnosing and
CC   monitoring the presence and metastasis of ovarian cancer and breast
CC   cancer. Amino acids ABJ01055 - ABJ01155 represent the ovary cell specific
CC   protein sequences of the invention
XX
SQ   Sequence 449 AA;

Alignment Scores:
Pred. No.:              3.61e-37      Length:         449
Score:                  400.00        Matches:        103
Percent Similarity:     58.1%         Conservative:   26
Best Local Similarity:  46.4%         Mismatches:     62
Query Match:            9.0%          Indels:         31
DB:                     5             Gaps:           7

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ABJ01057 (1-449)

Qy        588 GTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACC 647
              ::::::         ::::::|||         |||                ||||||
Db        243 LeuIleAsnAlaLeuIleTyrPhe---------LeuArgGlnSerLeuArgSerValAla 259

Qy        648 CAGGCATGAGCACAGTGGTGCAACCATAGG---------TCACTGCAGCCTCAACCTCCT 698
              ||||||      |||||||||:::                |||||||||||   ||||||
Db        260 GlnAlaGlyValGlnTrpCysSerGlyAlaAspLeuGlySerLeuGlnProLeuProPro 279

Qy        699 GAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCA 758
                    |||             |||||||||   ||||||||||||||          |||
Db        280 GlyPheLysAlaPheProCysLeuSerLeuLeuSerSerTrpAspTyrArgSerLeuPro 299

Qy        759 CCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAG 818
              |||      ||||||         |||   :::|||||||||   :::::::::   |||
Db        300 ProCysProAlaAsnPhePheValPheLeuIleGluThrGlyPheHisHisIleSerGln 319

Qy        819 GCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGG 878
                    :::   :::   |||      :::||||||      ||||||||||||||||
Db        320 ------IleSerIleSerAlaProCysAspProProAlaSerAlaSerGlnSerAlaGly 337
```

```
Qy        879 ATTACAGGGGTGAGCCACCATGTGCG---------GCTACTTATTTCTTTACATTCCATC 929
              |||||||||:::||||||  ||||||             |||:::  :::  ::::
Db        338 IleThrGlyMetSerHis--CysAlaGlnProAspValTyrTyrTyrValSerGlyTyrI 357

Qy        930 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 989
              ::                    |||:::|||  |||   :::|||     ||||||
Db        357 leGlyLys-------------GlnAspArgCysTyr-LeuPhePhePhePhePhe 371

Qy        990 TTTGAGACAGAGTCTCACTTCATCACC--------------------TCAACCTCCGTT 1028
              ||||||||||||||     :::                  |||   ::::::
Db        372 PheGluThrGluSerArgThrValAlaGlnAlaGlyArgLeuGluArgSerGlyAlaIle 391

Qy       1029 CAGC---------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
               |||         |||||||||||||||||||||||||   |||   |||||||||||||
Db        392 -SerThrArgArgSerLeuGlnProLeuProProGlyLeuLysArgPheSerCysLeuSe 411

Qy       1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
              |||||||||||||||:::|||:::      ||||||   ||||||::: ||||||   |
Db        411 rLeuLeuSerSerTrpAspTyrArgCysThrProProArgLeuAlaHis-PheCysThrP 431

Qy       1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195
              |||||||||||||||   ||||||      |||||||||   :::|||||||||
Db        431 heSerArgAspGlyValSerProCysTrpSerGlyTrpSerLeuSerProAspLeu 449

RESULT 14
AAU30391
ID   AAU30391 standard; protein; 304 AA.
XX
AC   AAU30391;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #882.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 287; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 304 AA;

Alignment Scores:
Pred. No.:              5.98e-37        Length:         304
Score:                  397.50          Matches:        123
Percent Similarity:     49.5%           Conservative:   22
Best Local Similarity:  42.0%           Mismatches:     62
Query Match:            9.0%            Indels:         87
DB:                     4               Gaps:           7

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAU30391 (1-304)

Qy        634 CTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAAC 693
              ::::::||||||   |||   |||   :::||||||   :::   ||||||:::|||:::
Db          3 ValSerLeuSerLeuArgLeuGluCysAsnGlyAlaIleLeuProHisCysAsnLeuHis 22

Qy        694 CTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTG 753
              |||       |||   |||          ||||||||||||||||||||   |||   |||
Db         23 LeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyIleThrGlyVal 42

Qy        754 CACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTG 813
              |||||||||      |||      :::      ||||||||||||   ::::::||||
Db         43 CysHisHisAlaProProIlePheValLeu--LeuValGluThrGlyPheHisHisValG 62

Qy        814 GCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTG 873
              ||||||||   |||:::||||||   ||||||:::   |||       ||||||||||||
Db         62 lyGlnAlaGlyLeuGluLeuLeuThrSerSerAspLeuProAlaTrpAlaSerGlnSerA 82

Qy        874 CTGGGATTACAGGGGTGAGCCAC--------------------------------CATG 900
              |||||||||||||   ||||||                                 |||
Db         82 laGlyIleThrGlyGlySerHisLeuCysLeuAlaAsnValLysTyrPheLysSerHisP 102

Qy        901 TGCGGCTA-----------------------------------CTTATTTCTTTACAT- 923
                :::::                                    |||:::||||||
Db        102 heLysIlePheValIleAspGluThrTrpPheGlnHisThr***LeuLeuSerLeuSerS 122

Qy        924 --------------------TCCATCTTTCCA-----------------ATAGAATGTA 945
                                  ||||||  |||                 :::   |||
Db        122 erCysAsp***Val***ThrSerIle***ProGlnValLeuAsp***AsnValLeuCysP 142

Qy        946 AGATCCACAGAACAGGGATTACTGCCT------------------------------- 972
                  |||   |||||||||
Db        142 ro-AspValSerGln***LeuLeuProTrpLeuIleLysIlePheLeuLysTyrThrLeu 161

Qy        972 ------------------------------------------------------------ 972

Db        162 AspCysTrp***LysArgGlnAlaTrpLeuTyrTrpTyr***PheGlyThrGluLeuPhe 181

Qy        973 -------------------------ATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTC 1005
                                        ||||||      |||||||||:::||||||
Db        182 GlyCysPro***GlyArgGlnThrLeuCysPhePhe------PheLeuArgArgSerLeu 199

Qy       1006 ACTTCATCACCTCAACCT--------------CCGTTCAGCTCACTGCAACCTCTGCCT 1050
                |||      |||                 ||||||||||||||||||   |||
Db        200 AspSerValAlaGlnAlaGlyValGln***CysAspLeuGlySerLeuGlnProLeuPro 219

Qy       1051 CCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCA 1110
              |||   |||||||||||||||||||||      :::||||||:::|||:::   |||
Db        220 ProArgPheLys***PheSerCysLeuSerPheProAsnSerTrpAspTyrArgCysAla 239

Qy       1111 CCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCC 1170
              ||||||      |||||||||      |||   :::      ||||||   ||||||
Db        240 ProProLeuProAlaAsnPheCysIlePheSerArgAspMetGlyPheAlaMetLeuVal 259

Qy       1171 AGGCTGGTCTCAAA-CTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 1229
              |||||||||||   ||||||||||   |||   |||||||   ||||||||||||
Db        260 ArgLeuValSerGluLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSerAla 279
```

```
Qy       1230 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
              ||||||   |||:::|||:::   |||||||||
Db        280 GlyIleThrGlyMetSerTyrArgAlaTrpPro 290

RESULT 15
AAB01399
ID   AAB01399 standard; protein; 375 AA.
XX
AC   AAB01399;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Neuron-associated protein.
XX
KW   Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW   ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW   Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
KW   Parkinson's disease; demyelinating disease; meningitis; prion disease;
KW   kuru; Creutzfeldt-Jakob disease; neurofibromatosis; cerebral palsy;
KW   muscular dystrophy; central nervous system; CNS;
KW   peripheral nervous system; PNS; myopathy; schizophrenia;
KW   actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW   myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW   autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW   AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
KW   ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW   Werner syndrome, trauma; human.
XX
OS   Homo sapiens.
XX
PN   WO200034477-A2.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US030408.
XX
PR   11-DEC-1998;   98US-00210083.
PR   09-FEB-1999;   99US-0119365P.
PR   16-MAR-1999;   99US-0124687P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Hillman JL,  Lal P,  Au-Young J;
PI   Yang J,  Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-423423/36.
XX
PT   New human neuron-associated proteins and polynucleotides encoding them,
PT   useful for diagnosis, treatment and prevention of cell proliferative
PT   disorders including cancer, neuronal and neurological disorders.
XX
PS   Disclosure; Page 143-144; 145pp; English.
XX
CC   Human neuron-associated proteins (NEUAP) can be used for for treating or
CC   preventing a disorder associated with decreased expression or activity of
CC   NEUAP. Antagonists of NEUAP are useful for treating or preventing
CC   disorder associated with increased expression or activity of NEUAP. NEUAP
CC   or their fragments or derivatives are useful for treating neurological
CC   disorder such as epilepsy, ischemic cerebrovascular disease, stroke,
CC   cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's
CC   disease, dementia and Parkinson's disease. NEUAPs are also useful for
CC   treating other demyelinating diseases, bacterial and viral meningitis,
CC   prion diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC   metabolic diseases of the nervous system, neurofibromatosis, other
CC   developmental disorders of the central nervous system, cerebral palsy,
CC   neuroskeletal disorders, autonomic nervous system disorders, cranial
CC   nerve disorders, spinal cord diseases, muscular dystrophy and other
CC   neuromuscular disorders, peripheral nervous system disorders, inherited,
CC   metabolic, endocrine, and toxic myopathies, mental disorders including
CC   mood, anxiety and schizophrenic disorders, a cell proliferative disorder
CC   such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
CC   cirrhosis, hepatitis, mixed connective tissue disease (MCTD),
CC   myelofibrosis, paroxysmal nocturnal haemoglobinuria, cancers of the
CC   adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, and an
CC   autoimmune/inflammatory disorder such as acquired immunodeficiency
CC   syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
CC   allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, Werner
CC   syndrome, complications of cancer, hemodialysis, and extracorporeal
CC   circulation, viral, bacterial, fungal parasitic, protozoal, and
CC   helminthic infections, and trauma. This protein was designated g3002527
XX
SQ   Sequence 375 AA;

Alignment Scores:
Pred. No.:              6.58e-37      Length:          375
Score:                  397.50        Matches:         105
Percent Similarity:     56.0%         Conservative:    12
Best Local Similarity:  50.2%         Mismatches:      60
Query Match:            9.0%          Indels:          32
DB:                     3             Gaps:            4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x AAB01399 (1-375)

Qy        597 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 656
              ::::::::::|||            |||                |||||||||||||||
Db        194 IlePheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
                ||||||   |||      ||||||||||||   ||||||      |||      :::
Db        214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSer 233

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
                  ||||||   ||||||||||||||   |||   ||||||||||||||||||
Db        234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
                  |||   |||||| :::|||   |||      |||||||||:::|||   |||
Db        253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuIleLeuIleSer-- 271

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
                      :::   |||      ||||||||||||||||||||||||||||||||||
Db        272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 291

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
              |||   ||||||:::   :::                          |||
Db        292 HisAlaArgLeuIlePheAsnPhe--------------------Cys------------ 300

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
                                                  ||||||   |||||||||   :::|||
Db        301 ----------------------------LeuPheGluMetGluSerHisSerValThr 310

Qy       1017 TCAA--------------CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 1061
               |||              |||         ||||||||||||||||||||||||   ||
Db        311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLy 330

Qy       1062 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 1121
              |   ||||||||||||||||   |||||||||:::|||        ||||||
Db        330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisPr 350

Qy       1122 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 1181
               |||||| ||||||||||||   |||   ||||||   |||         ||||||||||
Db        350 oAlaAsn-PheCysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerG 370

Qy       1182 AAACTCCTGACCTCAAG 1198
              |||||||||||||:::
Db        370 lnThrProAspLeuArg 375


Search completed: February  8, 2006, 10:43:53
Job time : 422.846 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:44:07 ; Search time 17.0675 Seconds
(without alignments)
2819.838 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 4426
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092020_27698/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_92_@runat_08022006_092020_27698 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	416	9.4	673	4	F40201	artifact-warning s
c	2	402.5	9.1	613	4	C40201	artifact-warning s
c	3	391.5	8.8	673	4	F40201	artifact-warning s
	4	391	8.8	613	4	C40201	artifact-warning s
c	5	338	7.6	627	4	A40201	artifact-warning s
	6	332.5	7.5	627	4	A40201	artifact-warning s
c	7	276.5	6.2	597	4	E40201	artifact-warning s
	8	269.5	6.1	418	2	S41044	chromosomal protei
	9	266.5	6.0	597	4	E40201	artifact-warning s
	10	264	6.0	196	2	I38022	hypothetical prote
	11	258	5.8	148	1	PSHU	phospholipase A2 (
c	12	248.5	5.6	301	4	B40201	artifact-warning s
	13	239.5	5.4	301	4	B40201	artifact-warning s
	14	239	5.4	579	4	D40201	artifact-warning s
c	15	232.5	5.3	100	2	A46010	X-linked retinopat
c	16	223.5	5.0	579	4	D40201	artifact-warning s
	17	218	4.9	146	1	PSDG	phospholipase A2 (
	18	216	4.9	144	2	JN0480	phospholipase A2 (
	19	216	4.9	146	1	PSRT	phospholipase A2 (
c	20	214.5	4.8	499	2	S65657	alpha-1C-adrenergi
	21	209	4.7	145	1	PSBOA	phospholipase A2 (
c	22	206.5	4.7	418	2	S41044	chromosomal protei
c	23	206	4.7	196	2	I38022	hypothetical prote
	24	205	4.6	146	1	PSPGA	phospholipase A2 (
	25	202	4.6	124	1	PSPGA2	phospholipase A2 (
	26	201.5	4.6	132	1	PSHOA	phospholipase A2 (
	27	188	4.2	146	1	S34049	phospholipase A2 (
c	28	172	3.9	46	2	I54375	gene NF2 protein -
c	29	162	3.7	39	2	I54374	gene NF2 protein -
c	30	160	3.6	369	2	A53959	thromboxane A-2 re
	31	159	3.6	119	2	S00227	phospholipase A2 (
	32	157	3.5	125	2	S65624	phospholipase A(2)
c	33	156	3.5	53	2	A42442	integrin beta-1 ch
c	34	155.5	3.5	407	2	T02670	probable thromboxa
c	35	154.5	3.5	79	2	A56194	thromboxane A-2 re
	36	149.5	3.4	119	1	PSNJ3K	phospholipase A2 (
c	37	147	3.3	841	1	I78885	serine/threonine-s
	38	144.5	3.3	146	1	PSNJAF	phospholipase A2 (
	39	144.5	3.3	499	2	S65657	alpha-1C-adrenergi
	40	143.5	3.2	119	1	PSNJ2K	phospholipase A2 (
c	41	143.5	3.2	124	2	A47582	B-cell growth fact
	42	143	3.2	133	1	PSOXG	phospholipase A2 (
	43	142.5	3.2	100	2	A46010	X-linked retinopat
	44	140.5	3.2	119	1	PSRIA	phospholipase A2 (
	45	139.5	3.2	118	2	S01801	phospholipase A2 (

ALIGNMENTS

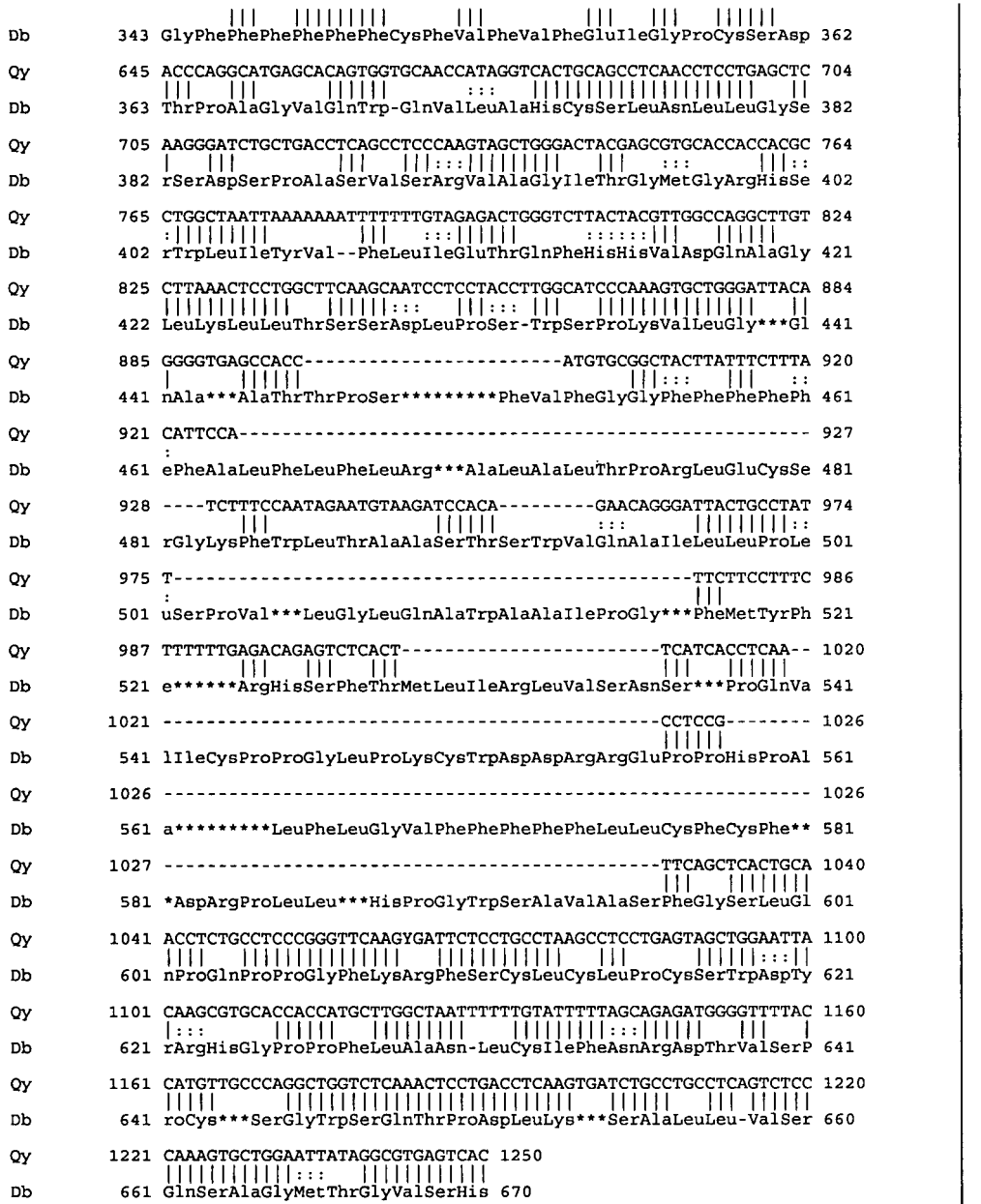
RESULT 1
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:
Pred. No.: 3.38e-35 Length: 673
Score: 416.00 Matches: 131
Percent Similarity: 41.8% Conservative: 17
Best Local Similarity: 37.0% Mismatches: 85
Query Match: 9.4% Indels: 122
DB: 4 Gaps: 9

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x F40201 (1-673)

```
Qy      532 AAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGG 591
            ||||||          ||||||               |||           |||   |||
Db      328 LysLysLysAsnProGlnLys--------------Gln*********PheCysPheTrp 342

Qy      592 TGACAGTTTGGGTTTTTTTTGTTGTTGTTTA-------GAGACAGGGTCTTGCTCTGTC 644
```

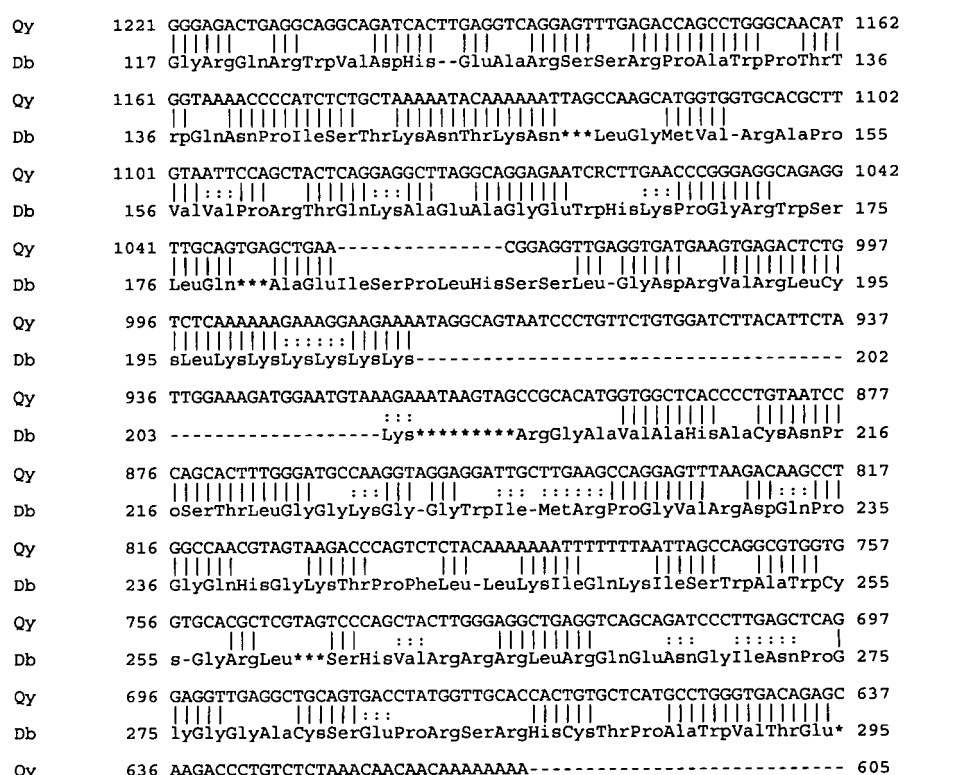


RESULT 2
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	9.08e-34	Length:	613
Score:	402.50	Matches:	162
Percent Similarity:	40.9%	Conservative:	48
Best Local Similarity:	31.6%	Mismatches:	151
Query Match:	9.1%	Indels:	152
DB:	4	Gaps:	17

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x C40201 (1-613)



```
Db        295 **AspSerValSerLysLysLysLysLys*********PhePhePhePhePhePheP 315

Qy        604 ----------------------ACCCAAACTGTCACCACTTACAAGCTGTGTGATCTTG 568
                                  ||||||      ::::::      |||||||
Db        315 heGluThrGluSerHisSerValThrGlnAlaGlyValGlnTrpArg------AspLeuG 333

Qy        567 GGCAAATAGCTCAGCCTTCTGAGCCTCAGTTTTCTTATCTGTTAAGTGAGGATAATTGTA 508
              ||   :::      |||      |||   |||
Db        333 lySerLeuGlnAlaPro---ProProGlyPhe---------------------------- 342

Qy        507 CCTATCCCATAGTGTTGTTCTGAGGATTAAAGGAGACACTGCCCAGAACCAAGAAAATTA 448
                                                :::|||   :::
Db        343 ------------------------------------MetProPheSerCysLeuSerL 350

Qy        447 ACACTAAATCATGG--------------------------CTGTTGTTACTATTATTTC 415
                  ::::::|||                         |||:::      :::|||
Db        350 euLeuArgThrTrpAspTyrArgArgProHisHisAlaGlnLeuIlePheCysIlePheS 370

Qy        414 CC----------------------CCCGGCCTACTGAGAACCAACTAGAATTCATAGG 379
                                     ||||||      ||||||
Db        370 erArgAsnGlyValLeuProCysTrpProGlyTrpSerArgThrPro------------ 385

Qy        378 TCAAGGAAGGGATAAACCTACTGCTACAGGTGATTGCCGAGCCAGAGCACGAGTATGAAT 319
                          |||::::::                |||:::
Db        386 -------------GlyLeuMetIleHisProProLeuProProLys-ValLeuGlyLeu 400

Qy        318 AGGTGTGGGTGTACGGKK---------------------------------------- 301
              :::      |||
Db        401 GlnAla***AlaThrAlaProArgPhePhePhePhePhe*********LeuArgGln 420

Qy        300 --------------------------TGTCCAGCAGAAATTTACAGC---TGTCCAGCT 271
                                       |||      ||||||   :::   |||
Db        421 SerLeuThrLeuSerProArgLeuGluCysAsnGlyGluIleSerAlaHisCysLysLeu 440

Qy        270 TCTTGGCCTGGTCRTAGCAGTTGTCATGTGTCTGGCAGCACCTGGAAAGTGGGAGGGACA 211
                |||||      |||||      |||   :::   |||   |||||
Db        441 HisLeuProGlyLeu------CysHisSerProAlaSerAlaPheCysValArgGlyThr 458

Qy        210 ------------------------------------------GCTGAGATAGGAGTAAGT 193
                                                       |||:::|||      |
Db        459 ThrGlyAlaArgThrMetProSer***PhePheValPheLeuValGluMetGlyPhe--C 478

Qy        192 GCAGAGCAATAGGTCAGCCCTCA---CCTGCCCACTCTCAGGAACAGGTGGGGATGACTT 136
              ||      :::|||||              :::              :::
Db        478 ysHis---ValGlyGlnAlaGlyLeuGluLeuLeuAlaSer***SerThrHisLeuCysL 497

Qy        135 CGCCAAGATGCTTC---AGGAGAAATGATCCTATGGCTTGAAAAAATATAAGTCCTTTTA 79
                |||:::|||:::      |||   :::|||:::                    |||
Db        497 euProLysCysTrpAspTyrArgArgGluProLeuHisProAla*********PhePheP 517

Qy         78 TCATATATTTATTGAAAGCATTTTTTTTTTTTGAGACAGAGTTGCTCTGTTACCCAGGCT 19
                                |||||||||   :::         ||||||:::|||||||
Db        517 he-----------------PhePhePhe***AspArgValSerLeuCysHisProGlyT 531

Qy         18 GGACAGCAGTGACACGA 2
              ||:::|||:::   |||
Db        531 rpSerAlaMetAlaArg 536

RESULT 3
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:
Pred. No.:                1.35e-32        Length:         673
Score:                    391.50          Matches:        122
Percent Similarity:       57.9%           Conservative:   21
Best Local Similarity:    49.4%           Mismatches:     65
Query Match:              8.8%            Indels:         41
DB:                       4               Gaps:           5

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x F40201 (1-673)

Qy       1246 CTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAGG 1187
              |||||||||:::|||||||||||||||||||||| |||:::|||||||||||||||||||
Db          6 LeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHisLeuArgSerGl 25

Qy       1186 AGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAAT 1127
              ||||   |||||||||   |||||||||:::|||   ||||||||||||||    ||||
Db         25 yVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLysIleHis-LysL 45

Qy       1126 TAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAGA 1067
              |||||::::::||||||      |||      ||||||      |||  |||||||||
Db         45 euAlaArgAsnGlyGlyProCysLeu***SerGlnLeuHisGlyArgGlnArgGlnGluA 65

Qy       1066 ATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG------------------CTGAACG 1025
              ||   ||||||||||||||   ||||||||||||                   |||
Db         65 snArgLeuAsnProGlyGly***GlyCysSerGluProLysLeuAlaThrAlaLeuGlnP 85

Qy       1024 GAGGTTG-AGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGT 966
                |||   :::      |||:::   |||||||||||:::|||:::      ::::::
Db         85 roGlyCysGln------SerLysGlyLeuSerGlnLysGlnLysGlnSerLysLysLys- 102

Qy        965 AATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTA- 907
                                                                  ||||||
Db        103 ------------------------------------------------LysLysLysThrP 107

Qy        906 --------------------GCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTG 867
                                  |||      |||||||||   |||   ||||||||||||
Db        107 roLysAsnLys*********AlaGlyCysGlyGlySerArgLeuSerSerGlnHisPheG 127

Qy        866 GGATGCCAAGGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTA 807
              ||   ||| |||   :::|||          |||||||||:::|||||||||   |||:::
Db        127 lyArgPro-GlyGlyGlnIleThr***GlyGlnGluPheGluThrSerLeuIleAsnMet 146

Qy        806 GTAAGACCCAGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTC 747
              |||:::      ||||||   |||  :::   |||   ||||||:::      ||||||
Db        147 ValLysLeuCysLeuTyr***Lys--TyrIleAsn***ProGlyMetAlaAlaHisAlaC 166

Qy        746 GTAGTCCCAGCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGG 687
                ::::|||||||||   |||      |||:::::|||          ||||||||||||
Db        166 ysAsnProSerTyrThrGlyAspArgGlyArgArgIleAla***ThrGlnGluValGluA 186

Qy        686 CTGCAGTGACCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGT 627
              ||||||||:::          ||||||   |||      |||   |||||||||
Db        186 laAlaValSerGlnAsn-LeuProLeuHisSerSerArgGlyValArgAlaArgAlaTyr 205

Qy        626 CTCTA-------AACAACAACAAAAAAAAAACCCAAACTGTCACCACTTACAAGCTGTGTG 574
              |||         |||      |||||||||   |||      |||   |||
Db        206 LeuLysAsnLysAsnLysAlaLysLysLysLysLysLysProProLysThrLys****** 225

Qy        573 ATCTTGGGC 565
                |||||
Db        226 ***LeuGly 228

RESULT 4
C40201
artifact-warning sequence (translated ALU class C) - human
```

```
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.  .

Alignment Scores:
Pred. No.:              1.51e-32      Length:         613
Score:                  391.00        Matches:        153
Percent Similarity:     40.2%         Conservative:   25
Best Local Similarity:  34.5%         Mismatches:     131
Query Match:            8.8%          Indels:         134
DB:                     4             Gaps:           14

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x C40201 (1-613)

Qy          2 TCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATGC 61
              |||   ||||||   ||||||||||||||||||          :::   ||||||||
Db        284 SerArgHisCysThrProAlaTrpValThrGlu***AspSerValSerLysLysLys--- 302

Qy         62 TTTCAATAAATATATGATAAAAGGACTTATATTTTT---------------------- 97
                          |||:::   |||
Db        303 --------------LysLysLys*********PhePhePhePhePhePheGluThr 317

Qy         98 TCAAGCCATAGGATCATTTCTCCTGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTG 157
                ||||||   :::                 ||||||
Db        318 GluSerHisSerValThrGlnAlaGlyValGlnTrpArgAsp----------------- 331

Qy        158 AGAGTGGGCAGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCT 217
                                                  |||      :::   |||   |||
Db        332 ----------------------------------LeuGlySerLeuGlnAlaProPro 339

Qy        218 CCCACTTTCCAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGA 277
              |||   |||      :::|||
Db        340 ProGlyPhe------MetPro-Phe---------------------------------- 345

Qy        278 CAGCTGTAAATTTCTGCTGGACAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGG 337
               |||||||      ||||||                  |||:::   |||
Db        346 -SerCysLeuSerLeuLeu--------------ArgThrTrpAspTyr---------- 356

Qy        338 CTCGGCAATCACCTGTAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGT 397

Db        356 ------------------------------------------------------------ 356

Qy        398 TCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTT 457
                  ||||||                  |||          |||||||||
Db        357 ----ArgArgPro----------------------HisHisAlaGlnLeuIlePheCy 367

Qy        458 GGTTCTGGGCAGTGTCTCCTTTAATCCT------------CAGAACAACACTATGGGATA 505
               :::                    |||                :::   |||   |||
Db        367 sIlePheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerArgThrProGly-L 387

Qy        506 GGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTG 565
                 :::|||   |||      |||           |||
Db        387 euMetIleHisProProLeuProProLysValLeuGlyLeuGlnAla***Ala------- 404

Qy        566 CCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGA 625
                    ||||||   ::::::      |||   |||||||||                 |
Db        405 -----------ThrAlaProArgPhe------PhePhePhePhePhe*********LeuA 419
```

```
Qy        626 GACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCA 685
              |||||   |||   ||||||||||||   |||   :::|||                |||||
Db        419 rgGlnSerLeuThrLeuSerProArgLeuGluCysAsnGlyGluIleSerAlaHisCysL 439

Qy        686 GCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTA 745
                ||:::|||                          ||||||      |||   |||||||
Db        439 ysLeuHisLeuProGlyLeuCysHisSerProAlaSerAlaPheCysValArgGlyThrT 459

Qy        746 CGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTA 805
              ||          |||   |||                  |||   ||||||   |||
Db        459 hrGlyAlaArg-ThrMetProSer***PhePheVal-PheLeuValGluMetGlyPheCy 478

Qy        806 CTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATC 865
               :::||||||||||||   |||:::||||||||||||     |||::::::|||   :::
Db        478 sHisValGlyGlnAlaGlyLeuGluLeuLeuAlaSer***--SerThrHisLeuCysLeu 497

Qy        866 CCAAAGTGCTGGGATTACAGGGGTGAGCCA------CCATGTGCGGCTACTTATTTCTTT 919
              ||||||||||||||||||||   ||||||      |||           :::||||||
Db        498 ProLysCysTrpAspTyrArgArgGluProLeuHisProAla*********PhePhePhe 517

Qy        920 ACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCT 979

Db        517 ------------------------------------------------------------ 517

Qy        980 TCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCAC---------------CTCAACCTC 1024
                |||   |||   ||||||||||||||   |||                 :::
Db        518 ---PhePhePhe***AspArgValSerLeuCysHisProGlyTrpSerAlaMet-AlaAr 536

Qy       1025 CGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCC 1084
                 :::|||||||||:::|||   ||||||||      ||||||||||||:::|||
Db        536 gSerArgLeuThrAlaSerSerThrSerArgValTyrAlaIleLeuLeuProGlnProSe 556

Qy       1085 TGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGC 1144
                    :::   |||||||||          ||||||   ||||||||||||||||||
Db        556 rAlaTyrValGlyLeuGlnAla-ProAlaProCysProAlaAsnPheLeuTyrPhe**** 576

Qy       1145 AGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCT 1204
                :::|||||||||   ||||||   ||||||||||||||||||   |||      |||
Db        576 **LysTrpGlyPheAlaMetLeuAlaArgLeuValSerAsnSerTrpProHis--AspPr 595

Qy       1205 GCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGT 1253
                |||         ||||||||||||||||||   ||||||||||||||
Db        595 oProThrPheAlaSerGlnSerAlaGlyIleThrGlyValSerHisCys 611

RESULT 5
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              6.37e-27      Length:         627
Score:                  338.00        Matches:        144
Percent Similarity:     43.0%         Conservative:   38
Best Local Similarity:  34.0%         Mismatches:     129
```

```
Query Match:          7.6%      Indels:       112
DB:                   4         Gaps:         12

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A40201 (1-627)

Qy      1247 ACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGATCACTTGAGGTCAG 1188
Db         5 SerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyGlyArgIleThr--ArgSerG 24

Qy      1187 GAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTAAAAATACAAAAAA 1128
Db        24 lyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleGlnLysI 44

Qy      1127 TTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAGGCTTAGGCAGGAG 1068
Db        44 leSerArgAla-***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu 63

Qy      1067 AATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG--------------CTGAACGGA 1023
Db        64 AsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys-ThrPr 83

Qy      1022 GGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAAT 963
Db        83 oAlaLeuAlaThrGluArgAspSerValSerGluLys---------------------- 95

Qy       962 CCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAGTAGCCG 903
Db        96 ----------------------------AsnLysAsnLysLys***Lys*********Al 106

Qy       902 CACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGATTGCT 843
Db       106 aGlyArgGlyLeuThrProValIleProAlaLeuTrpGluAlaLysAlaGlyGly--Ser 125

Qy       842 TGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAAAAAA 783
Db       126 ArgGlyGlnGluIleGluThrIleLeuAlaThrThrValLysProArgLeuTyr-***Ly 145

Qy       782 ATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGAGGCT 723
Db       145 sTyrLysLysLeuAlaGlyArgSerAlaGly-AlaCysSerProSerTyrLeuGlyGly* 165

Qy       722 GAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCACCA 663
Db       165 **GlyArgArgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaT 185

Qy       662 CTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAACAACAAAAAAAAAC 603
Db       185 hrAlaLeuGlnProTrpArgGlnSerGluThrProSerArgLysLysIleLysThrLysA 205

Qy       602 CCAAA--------------------CTGTCACCACTTACAAGCTGTGTGATCTTGGGCA 564
Db       205 snLysLys*********ProGlyAlaValSerArgLeu***Ser---------------- 219

Qy       563 AATAGCTCAGCCTTCTGAGCCTCAGTTTTCTTATCTGTTAAGTGAGGATAATTGTACCTA 504
Db       219 ------------------------------------------------------------ 219

Qy       503 TCCCATAGTGTTGTTCTGAGGATTAAAGGAGACACTGCCCAGAACCAAGAAAATTAACAC 444
Db       220 ----------------------------GlnHisPheGlyArgProArgArg------A 228

Qy       443 TAAATCATGGCTGTTGTTACTATTATTTCCCCCCGGCCTACTGAGAACCAACTAGAATTC 384
Db       228 laAspHis----------------------------------------------GluValA 233

Qy       383 ATAGGTCAAGGAAGGGATAAACCTACTGCTACAGGTGATTGCCG---AGCCAGAGCACGA 327
Db       233 rgArgSerArgPro------SerTrpLeuProArg***AsnProValSerThrLysAsnT 251

Qy       326 GTATGAATAGGTGTGGGTGTACGGKKTGTCCAGCAGAAATTTACAGCTGTCCAGCTTCTT 267
Db       251 hrLysAsn***ProGlyValValArgAlaProVal------------ValProAlaThrT 267

Qy       266 GGCCTGGTCRTAGCAGTTGTCATGTGTCTGGCAGCACCTGGAAAGTGGGAGGGACAGCTG 207
Db       267 rp------------------------Glu-AlaGluAlaGlyGluTrpArgGluProGly 278

Qy       206 AGATAGGAGTAAGTGCAGAGCAATAGGTCAGCCCTCACCTGCCCACTCTCAGGAACAGGT 147
Db       279 Arg------------ArgSerLeuGln***AlaGluIleAlaProLeuHisSerSerLeu 294

Qy       146 GGGGATGACTTCGCCAAGATGCTTCAGGAGAAATGATCCTATGGCTTGAAAAAATATAAG 87
Db       295 GlyAspArgAlaArgLeuArgLeuGlyLysLys***LysGlnLysIleLysLys****** 314

Qy        86 TCCTTTTATCATATATTTATTGAAAGCATTTTTTTTTTTTGAGACAGAG 38
Db       315 ***PhePheTyr-PheLeuPhe------LeuPhePheSerGluThrGlu 328

RESULT 6
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:
Pred. No.:              2.44e-26      Length:         627
Score:                  332.50        Matches:        114
Percent Similarity:     46.6%         Conservative:   18
Best Local Similarity:  40.3%         Mismatches:     86
Query Match:            7.5%          Indels:         65
DB:                     4             Gaps:           4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A40201 (1-627)

Qy       597 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 656
Db       320 LeuPheLeuPhePheSer---------GluThrGluSerArgSerValAlaLysAlaGly 336

Qy       657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
Db       337 ValGlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyPheThrProPheSer 356

Qy       717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
Db       357 CysLeuSerLeuProSerSerTrpAspTyrArgArgPro-HisTyrAlaArgLeuIlePh 376

Qy       777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
Db       376 eCysIle-PheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSerProA 396

Qy       837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
Db       396 spLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGlnAla-------- 412

Qy       897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
Db       413 ------------------------------------------***AspArgAlaArg** 418
```

```
Qy       957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
                                  |||      |||   |||:::||||||      |||||
Db       418 *******PhePheIlePheCysPheTyrPhePheProArgArgSerLeuAlaLeuSerPr 438

Qy      1017 TCAACCTCCGT---------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 1061
             |::: |||                |||||||||||   |||   ||||||||||||
Db       438 oArg-LeuGluCysSerGlyAlaIleSerAlaHisCysLysLeuArgLeuProGlySerA 458

Qy      1062 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 1121
                |||||||||   ||||||   |||||||||   |||           |||
Db       458 rgHisSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyAlaArg-ThrThrPro 477

Qy      1122 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 1181
             |||   |||||||||||||||   |||   ||||||:::      :::|||   ||||||
Db       478 Gly***PhePheValPheLeuValGluThrGlyPheHisArgGlySerGlnAspGlyLeu 497

Qy      1182 AAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGC 1241
                |||||||||||    ||||||   |||   |||||||||||||||:::|||||||
Db       498 AspLeuLeuThrSer***--SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgA 517

Qy      1242 GTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCT 1301
             |||||  |||   ||||||           :::||||||:::
Db       517 rgGlu--ThrAlaProGly*********PheLeuPhePheVal----------------- 530

Qy      1302 AGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGA 1361

Db       530 ------------------------------------------------------------ 530

Qy      1362 TATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGA 1410
                        |||:::|||   :::            |||||||||
Db       531 ------------PheIlePhePheArgAspGlyValSerLeuCysArg 542

RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              2.13e-20        Length:         597
Score:                  276.50          Matches:        154
Percent Similarity:     37.2%           Conservative:   35
Best Local Similarity:  30.3%           Mismatches:     136
Query Match:            6.2%            Indels:         186
DB:                     4               Gaps:           18

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x E40201 (1-597)

Qy      1261 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1202
             ||||||          |||   |||:::|||||||||||||||||   :::|||||||||
Db          1 GlyGlnValTrp***LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArg 20

Qy      1201 TC-ACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGC 1143
                |||   |||   |||||||||||||||||   ||||||   ||||||||||||
Db         21 IleThr***GlyTrpGluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr 40

Qy      1142 TAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGG 1083
             ||||||  |||||||||   |||||||||          |||||||||||||||:::|
Db         41 ***LysTyr--LysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgG 60

Qy      1082 AGGCTTAGGCAG---GAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTG--- 1029
             |||||          ||||||   ||||||||||||   :::   |||   |||:::
Db         60 luAla-GluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGluIleVal 79

Qy      1028 ------AACGGAGGTTGAGGTGATGA-AGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAA 976
                    |||   |||:::   |||:::|||   |||||||||:::|||
Db         80 ProSerTyrSerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys****** 99

Qy       975 AATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAG 916
                 :::             |||||||||
Db       100 ***AlaArg------------CysGlySer------------------------------ 105

Qy       915 AAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGG 856
                                 |||   |||   ||||||   ||||||      |||
Db       106 ----------------------SerCysLeu***SerGlnArgPheGlySerSerArg 117

Qy       855 TA-GGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 797
                ||||||      |||      |||||||||   |||   ||||||      :::|||
Db       118 ArgGlyGlySerProGluValGlySerLeuArgProAla***ProThrTrpArgAsnPro 137

Qy       796 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 737
             :::|||||||||                |||||   ||||||||||||      ||||||
Db       138 IleSerThrLysAsnIleLys---LeuAlaGlyArgGlyGlyAlaCysLeu***SerGln 156

Qy       736 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 677
             ||||||||||||||||      :::   ::::::
Db       157 LeuLeuGlyArgLeuArgGln---GluArgIleAla***ThrGlnGluAlaLysPheAla 175

Qy       676 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAA------------------ 635
                :::   |||   |||         ||||||   |||
Db       176 ValArgLeuCysHisArgThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys 195

Qy       634 ---------------------------------------GACCCT--------------- 629
                                                       :::|||
Db       196 LysLys*********ProGlyValValAlaHisAlaCysAsnProSerAlaLeuGlyVal 215

Qy       628 ----------------------------GTCTCTAAACAACAACAAAAAAAAAACCCAA 599
                                         |||      |||      :::      :::
Db       216 GlnGlyGlyAlaAspHisLeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGlu 235

Qy       598 ACTGTCACCACTTACAAGCTGTGTGAT---CTTGGGCAAATAGCTCAGCCTTCTGAGCCT 542
                |||   :::      |||:::   :::   ||||||   :::|||              |||
Db       236 ThrProSerLeuLeuLysIle***Asn***LeuGlyValValAlaHisAlaCysAsnPro 255

Qy       541 CAGTTTTCTTATCTGTTAAGTGAGGATAATTGTACCTATCCCATAGTGTTGTTCTGAGGA 482
                |||:::
Db       256 -SerTyr------------------------------------------SerGlyGly** 261

Qy       481 TTAAAGGAGACACTGCCCAGAACCAAGAAAATTAACACTAAATCATGGCTGTTGTTACTA 422
                |||||          ||||||||||   :::|||          |||      :::
Db       261 *AspArgArgGluLeuLeuGluProArgLysGlnSerLeuArg***AspCysAla--Ile 280

Qy       421 TTATTTCCCCCCGGCCTACTGAGAACCAACTAGAATTCATAGGTCAAG------------ 374
                :::      ||||||           |||      |||      |||
Db       281 ValLeuArgProGlyGlnGlnGluGlnAsnSerValSerLysLysLysLys********* 300

Qy       373 --------------GAAGGGATAAACCTACTGCTACAGGTGATTGCCGAGCCAGAGCAC 329
                           :::|||:::   |||||||||              |||
Db       301 PhePhePhePhe***AspGlyValLeuLeuLeuLeu--------------ProArgPro 315

Qy       328 GAGTATGAATAGGTGTGGGTGTACGGKKTGTCCAGCAGAAATTTACAGCTGTCCAGCTTC 269
             ||||||:::           |||      |||              |||      |||
Db       316 GluTyrAsp--------------GlyThrIleSerProGlnThrLeuLeuProGlyPhe 330

Qy       268 TTGGCCTGGTCRTAGCAGTTGTCATGTGTCTGGCAGCACCTGGAAAGTGGGAGGGACAGC 209

Db       330 ------------------------------------------------------------ 330

Qy       208 TGAGATAGGAGTAAGTGCAGAGCAATAGGTCAGCCCTCACCTGCCCACTCTCAGGAACAG 149
```

```
                                                  |||   ||||||      :::
Db        331 -------------------------LysGlnPheSerProValSerAlaSerArgVal 341

Qy        148 GTGGGGATGACTTCGCCAAGATGCTTCAGGAGAAATGATCCTATGGCTTGAAAAAATATA 89
                ||||:::|||      |||                                  :::
Db        342 AlaGlyIleThrGlyMetArg------------------------------HisHis 350

Qy         88 AGTCCTTTTATCATATAT-------------------------TTATTGAAA------ 62
              :::      |||:::|||                        |||:::|||
Db        351 AlaGlnLeuIleLeuTyrPhe******ArgTrpGlyPheSerMetLeuValLysLeuVal 370

Qy         62 ------------------------------------------------------------ 62

Db        371 SerAsnSerGlnProGlnValIleArgProAlaLeuAsnSerGlnSerAlaGlyIleThr 390

Qy         61 -------------------------GCATTTTTTTTTTTTGAGACAGAGTTGCTC--- 32
                                       ||||||||||||||||||||
Db        391 GlyMetSerTyrHisThrTrp********PhePhePhePheGluThrGluPheCysSer 410

Qy         31 TGTTACCCAGGC 20
              |||   ||||||
Db        411 CysCysProGly 414
```

RESULT 8
S41044
chromosomal protein - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
C;Accession: S41044
R;Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A;Title: A new chromosomal protein essential for mitotic spindle assembly.
A;Reference number: S41044; MUID:94166884; PMID:8121495
A;Accession: S41044
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-418 <YEO>
A;Cross-references: UNIPROT:P49646; UNIPARC:UPI000013C29A; GB:L26953; NID:g537529; PIDN:
C;Superfamily: very very hypothetical protein RMSA-1
C;Keywords: chromosomal protein

Alignment Scores:

Pred. No.:	1.14e-19	Length:	418
Score:	269.50	Matches:	122
Percent Similarity:	37.0%	Conservative:	51
Best Local Similarity:	26.1%	Mismatches:	107
Query Match:	6.1%	Indels:	188
DB:	2	Gaps:	20

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x S41044 (1-418)

```
Qy       1428 TTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTG 1487
              ||||||         :::::::|||         |||::::::|||         |||:::
Db          4 PheCysLeuIleLeuPheAsnLeu---------PheAsnPheCys---------LeuIle 17

Qy       1488 TTACTGTTTTTTCATTGT-----------------TACTATTTTAGT------------ 1517
              |||   |||      |||                 |||:::||||||
Db         18 LeuSerPheLeuThrCysIleAlaTyrProValLeuTyrPhePheSerPheLeuAsnSer 37

Qy       1518 --------------------------------CACTGAATAATATGG------------ 1532
                                               ||||||
Db         38 GlnThrArgSerLeuLysLeuPheLeuTyrPheLysIleGluIleTrpSerGlnGlyIle 57

Qy       1533 CTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG--- 1589
              :::      |||:::                 |||   :::||||||:::|||
Db         58 ValArgPheLeuLeuArgPheGlyLeuThrThrPheLysGluArgPheThrSerLeuIle 77

Qy       1589 ------------------------------------------------------------ 1589

Db         78 LeuLeuAsnAsnMetHisGlnMetIlePheProMetValLysTyrIleSerLysLeuCys 97

Qy       1590 -----------------ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACAC 1631
                              :::      |||:::   :::||||||:::      :::

Db         98 IlePheHisPheTrpHisLeuValLeuMetAspLeuValProArgGlnGlnArgSer-Il 117

Qy       1632 ACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTA 1691
                  |||   :::      |||   ::::::          |||   |||   ::::::
Db        117 eIleThrTyrSerLeuValPheAlaIleIleSerGlnLysLysLysArgGlyIleTyrHi 137

Qy       1692 GAGGATGTTTGTTCATATT---------------------------------------- 1710
                 :::      :::   |||
Db        137 sLysAsnAsnIleArgIleIleLeuPheLeuProGlnAlaHisGlyArgAspPheTyrVa 157

Qy       1711 ----------------AATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAAC 1754
                                         :::   |||   :::   |||
Db        157 lProIleLeuProPheThrGlnSerTyrValAspTrpGlyArgTrpLeu----------- 173

Qy       1755 CACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGC 1814
                    :::|||::::::|||   |||   ||||||||||||||||||||||||||||
Db        174 -------IleTrpGluAlaLysAlaGlyGluSerLeuGluValArgSerSerArgProAl 191

Qy       1815 CTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTG 1874
                |                |||   :::||||||||||||   |||||||||
Db        191 aSerGlnSerArgArgAsnSerValSerThrLysAsnIleLysIleSer----------- 207

Qy       1875 ATGCATGCCTGTA---------------------------------------------- 1887
                      ||||||
Db        208 -------ProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGl 225

Qy       1888 ------------GTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGG 1934
                         ||||||   ||||||:::|||||||||   ||||||   :::||| |
Db        225 yValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAspPro-G 245

Qy       1935 GAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGC 1994
              ||||||||||||||:::|||   ||||||   ||||||||||||||||||   |||
Db        245 lyGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThrSer- 264

Qy       1995 GAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAA 2054
                :::::::::|||||||||          |||   |||   :::::::::|||:::::::
Db        265 --GluThrValSerLysLys------LysLysLysLysLysLysLysLysThrArgValL 282

Qy       2055 TAAGATTCCTGAAGGTAAGCAGA-----------GATACGTAAATTATATGTAATAAA- 2101
                ::      :::      |||                  |||   |||   |||   |||
Db        282 euThrCysIleAsnAlaSerThrLeuPheHisValLeuThrArgIlePheCysTyrLysG 302

Qy       2102 --------------------GTTTAAATGCATTTTAACTGTAATCTTATTGTTTAT---- 2137
                                    :::              :::   ||||||
Db        302 lnLysTyrProLeuTrpLeuLeuIleLysArgLeuSerValAsnLeuGlnLysGlnGlnG 322

Qy       2138 -----TTTGGTTATAAAAGTAAACAAGCCAAAAGT---------AATGCAACTTCAAACT 2183
                        |||   ||||||   |||                  |||:::   :
Db        322 luAspLeuGlnArgLysLysLysGlnArgLysGlnGluHisProGluLeuThrAlaGlnA 342

Qy       2184 CTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAA 2243
                ::|||   :::   ::::::      |||      |||   :::   ||||||   :::|
Db        342 snThrValLeuLysMetLeuLeuGlnGly---ValLeuIleAsnLeuTyrProThrValT 361

Qy       2244 CCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAAT 2303
                ||
Db        361 hrLeu------------------------------------------------------- 362

Qy       2304 TTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATG 2363

Db        362 ------------------------------------------------------------ 362

Qy       2364 TTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTC 2423
                                                        |||   |||:::|||||||
Db        363 ------------------------------------AsnArgProGlyValValAlaH 370

Qy       2424 ACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTT 2483
              ||||||||||||||||||||   ||||||          ||||||      |||
Db        370 isAlaCysIleProAlaLeuTrp---AlaGluGlyGly-ValHisLeuAsnProGly--- 387

Qy       2484 TGAGACCAGCCTGGACAA 2501
                |||||||||||
Db        388 ------GlnProGlyGln 391
```

```
RESULT 9
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                2.45e-19      Length:          597
Score:                    266.50        Matches:         150
Percent Similarity:       37.2%         Conservative:    30
Best Local Similarity:    31.0%         Mismatches:      169
Query Match:              6.0%          Indels:          138
DB:                       4             Gaps:            21

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x E40201 (1-597)

Qy          5 TGTCACTGCTGTCCAGCCTGGGTAACAGAG---CAACTCTGTCTCAAAAAAAAAAAATGC 61
              ||||||     |||||||||   |||      :::|||   ||||||||||||||
Db        179 CysHisArgThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLysLysLys-** 198

Qy         62 TTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGA---------- 110
                      ::::::                     |||:::|||
Db        198 *******ProGlyValValAlaHisAlaCysAsnProSerAlaLeuGlyValGlnGlyGl 218

Qy        111 ----------------------------------------------------------TCATT 115

Db        218 yAlaAspHisLeuArgLeuGlyVal***AspGlnLeuAspGlnHisGlyGluThrProSe 238

Qy        116 TCTCCTGAAG---------CATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGC 166
               |||||||||         ||||||   |||        |||:::      :::::
Db        238 rLeuLeuLysIle***Asn***LeuGlyValValAlaHisAlaCysAsnProSerTyrSe 258

Qy        167 AGGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTC 226
               :|||                          ||||||                  |||
Db        258 rGly---------Gly***AspArgArgGluLeuLeuGluProArgLysGlnSerLeuAr 275

Qy        227 CAGGTGCTGC----------------------CAGACACATGACAACTGCTAYGACCAGGC 265
                    |||                     |||         |||          :::
Db        275 g***AspCysAlaIleValLeuArgProGlyGlnGlnGluGlnAsnSerValSerLysLy 295

Qy        266 CAAGAAG----------------------CTGGACAGCTGTAAATTTCTGCTGGACAMMCC 304
               |||||                     |||          ||||||      ||
Db        295 sLysLys*********PhePhePhePhe***AspGlyValLeuLeuLeuProArgPr 315

Qy        305 G---TACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAG 361
               |   |||      |||   |||
Db        315 oGluTyrAspGlyThrIleSer-----------------------------ProGlnTh 325

Qy        362 GTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAA 421
                  :::|||                  :::      ||||||      |||
Db        325 rLeuLeuProGlyPheLysGlnPheSerProValSerAlaSerArgValAlaGlyIleTh 345

Qy        422 TAGTAACAACAGCCAT---GATTTAGTGTTAATTTTC---------------------- 455
                        |||     |||:::|||   |||
Db        345 rGlyMetArgHisHisAlaGlnLeuIleLeuTyrPhe******ArgTrpGlyPheSerMe 365

Qy        456 ----------TTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATA 505
                       ||||||     |||    |||   ::::::|||       |||
Db        365 tLeuValLysLeuValSerAsnSer-GlnProGlnValIleArgProAlaLeu------- 382

Qy        506 GGTACAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTG 565
                                                        :::|||   |||   :::
Db        383 ----------------------------------AsnSerGlnSerAlaGlyIleThrG 391

Qy        566 CCCAAGAT---CACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTT 622
                      ||||||         |||                ||||||          |||
Db        391 lyMetSerTyrHisThr---------Trp-*********PhePhePhe---------Phe 404

Qy        623 AGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATA----GGT 678
                              |||   ||||||   ||||||:::   |||
Db        405 GluThrGluPheCysSerCysCysProGlyArgSerThrMetAlaGlnSerHisArgLys 424

Qy        679 CACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCT 738
               |||   |||         ||||||      ||||||::: |||||||||:::   ||
Db        425 LeuCysPheLeuGly------SerSerAsnSerLeuLeuSer-GlnProProGlu***Le 442

Qy        739 GGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTG 798
               |||||||:::||||||   |||||||||                  |||   ||||||
Db        442 uGlyLeuGlnAlaCysAlaThrThrProSer---***PheTyrIlePheSerArgAspGl 461

Qy        799 GGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTA-CC 857
               |||         |||         |||   ||||||      :::   |||:::   ||
Db        461 yValSerProCysTrpSerSerTrpSerGlnThrProAsnLeuArg***SerAlaProPr 481

Qy        858 TTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCT 917
               |      ||||||   ||||||||||||      ||||||      |||
Db        481 o***ThrProLysAlaLeuGlyLeuGlnAla***AlaThrThrProGly*********Ph 501

Qy        918 TTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTT 977
                :::
Db        501 ePhe---------------------------------------------------------- 502

Qy        978 CTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA----------------- 1020
                        |||||||||:::|||                |||
Db        503 ----------PheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValArgTrpHi 519

Qy       1021 ----------CCTCCGTTCAGCTCA---CTGCAACCTCTGCCTCCCGGGTTCAAGYGATT 1067
                           |||:::|||   :::|||   :::
Db        519 sAsnLeuThrAlaAsnPheAlaSerTrpValGlnAlaIle-----------------Le 533

Qy       1068 CTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA 1127
               |||||||||||||||   |||||||||:::|||:::   |||||||||      |||||
Db        533 uSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArgProAlaAs 553

Qy       1128 TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTC 1187
               |  |||:::||||||   ||||||||||   ||||||   ||||||||||||||||||
Db        553 n--PheIlePheLeuValGluMetGlyPheLeuHisValGlyGlnAlaGlyLeuLysLeu 572

Qy       1188 CTGACCTCAAGTGATCTGCCTGC-CTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAG 1246
               ||||||   |||   |||   |||   |||||||||   |||:::||||||   |||
Db        573 ProThrSerGlyAspProProArgLeuGluLeuProLysArgTrpAspTyrArgHisGlu 592

RESULT 10
I38022
hypothetical protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C;Accession: I38022
R;Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A;Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A;Reference number: I38021
A;Accession: I38022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI
```

```
Alignment Scores:
Pred. No.:                4.1e-19       Length:          196
Score:                    264.00        Matches:         61
Percent Similarity:       68.2%         Conservative:    12
Best Local Similarity:    57.0%         Mismatches:      23
Query Match:              6.0%          Indels:          12
DB:                       2             Gaps:            3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x I38022 (1-196)

Qy       1787 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1846
              |||||||||      |||:::|||   |||||||||:::|||:::      |||||||||
Db          3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22

Qy       1847 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1906
              |||||||||:::|||||||||               ||||||   ||||||||||||   |||
Db         23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42

Qy       1907 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1966
              |||||||||||||||  ||||||  |||||||||||||||||||||||:::||||||
Db         43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62

Qy       1967 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 2026
                 |||  |||||||||   |||:::   ||||||:::||||||:::      ||||||
Db         63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGln------AsnLys 80

Qy       2027 ATAAAGGATCGGAG-------AGAAACAAAACTAATAAGATT----------------- 2061
                 |||:::            ||||||   |||   |||
Db         81 ThrLysAsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysIleIleGly 100

Qy       2062 ---CCTGAAGGTAAGCAGAGA 2079
                 |||||||||:::::::::
Db        101 GlyProGluGlyGluGluLys 107

RESULT 11
PSHU
phospholipase A2 (EC 3.1.1.4) IB precursor [validated] - human
N;Alternate names: pancreatic phospholipase A2; phosphatidylcholine 2-acylhydrolase
C;Species: Homo sapiens (man)
C;Date: 17-Mar-1987 #sequence_revision 31-Mar-1993 #text_change 16-Aug-2004
C;Accession: C25793; A25793; A00733; A21566
R;Seilhamer, J.J.; Randall, T.L.; Yamanaka, M.; Johnson, L.K.
DNA 5, 519-527, 1986
A;Title: Pancreatic phospholipase A2: isolation of the human gene and cDNAs from porcine
A;Reference number: A25793; MUID:87132925; PMID:3028739
A;Accession: C25793
A;Molecule type: DNA
A;Residues: 1-148 <SEI2>
A;Cross-references: UNIPROT:P04054; UNIPARC:UPI0000048FD9; GB:M22970
A;Accession: A25793
A;Molecule type: mRNA
A;Residues: 1-148 <SEI>
A;Cross-references: UNIPARC:UPI0000048FD9; GB:M21055
A;Experimental source: lung
R;Grataroli, R.; Dijkman, R.; Dutilh, C.E.; Van der Ouderaa, F.; De Haas, G.H.; Figarell
Eur. J. Biochem. 122, 111-117, 1982
A;Title: Studies on prophospholipase A2 and its enzyme from human pancreatic juice. Cata
A;Reference number: A91114; MUID:82138816; PMID:7060561
A;Accession: A00733
A;Molecule type: protein
A;Residues: 16-22 <VER1>
A;Cross-references: UNIPARC:UPI000014C3BF
R;Verheij, H.M.; Westerman, J.; Sternby, B.; De Haas, G.H.
Biochim. Biophys. Acta 747, 93-99, 1983
A;Title: The complete primary structure of phospholipase A2 from human pancreas.
A;Reference number: A90651; MUID:83283533; PMID:6349696
A;Accession: A21566
A;Molecule type: protein
A;Residues: 23-143,'YS',146-147 <VER2>
A;Cross-references: UNIPARC:UPI0000172758
C;Genetics:
A;Gene: GDB:PLA2G1B; PLA2A; PLA2; P
A;Cross-references: GDB:120715; OMIM:172410
A;Map position: 12q23-12qter
A;Introns: 12/1; 65/2; 108/1
C;Function:
A;Description: catalyzes hydrolysis of 1,2-diacyl-sn-glycero-3-phosphocholine to 1-acyl
A;Note: the reaction is strongly enhanced when the phospholipid is condensed into a mic
C;Superfamily: Phospholipase A2
C;Keywords: calcium; carboxylic ester hydrolase; lipid degradation; lipid digestion; me
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-22/Domain: activation peptide #status experimental <ACT>
F;23-148/Product: phospholipase A2 IB #status experimental <MAT>
F;33-99,49-146,51-67,66-127,73-120,83-113,106-118/Disulfide bonds: #status predicted
F;50,52,54,71/Binding site: calcium (Tyr, Gly, Gly, Asp) #status predicted
F;70,121/Active site: His, Asp #status predicted

Alignment Scores:
Pred. No.:                1.73e-18      Length:          148
Score:                    258.00        Matches:         55
Percent Similarity:       59.8%         Conservative:    6
Best Local Similarity:    53.9%         Mismatches:      12
Query Match:              5.8%          Indels:          29
DB:                       1             Gaps:            3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x PSHU (1-148)

Qy        113 ATTTCTCCTGAAGCATCTTGGCGA--------------------------------AGT 139
              |||||||||   |||   |||:::                                  |||
Db         19 IleSerProArgAlaValTrpGlnPheArgLysMetIleLysCysValIleProGlySer 38

Qy        140 CATCCCCACCTG--------------------TTCCTGAGAGTGGGCAGGTGAGGGCTG 178
                 |||   |||                  :::      :::||| |||   |||
Db         39 AspProPheLeuGluTyrAsnAsnTyrGlyCysTyrCysGlyLeuGly-GlySerGlyTh 58

Qy        179 ACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCA 238
                 |||::::                                  |||   :::||||||||
Db         58 rProValAsp------------------------------GluLeuAspLysCysCysGl 68

Qy        239 GACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGA 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 nThrHisAspAsnCysTyrAspGlnAlaLysLysLeuAspSerCysLysPheLeuLeuAs 88

Qy        299 CAMMCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAG 358
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 pAsnProTyrThrHisThrTyrSerTyrSerCysSerGlySerAlaIleThrCysSerSe 108

Qy        359 TAGG 362
                 |:::
Db        108 rLys 109

RESULT 12
B40201
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC9
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:
Pred. No.:                1.88e-17      Length:          301
```

```
Score:                  248.50        Matches:        80
Percent Similarity:     43.7%         Conservative:   17
Best Local Similarity:  36.0%         Mismatches:     74
Query Match:            5.6%          Indels:         52
DB:                     4             Gaps:           8

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x B40201 (1-301)

Qy       1261 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1202
Db          1 GlyArgAlaTrpTrpLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGly 20

Qy       1201 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 1142
Db         21 LeuLeuGluThrArgSerSerArgProAlaTrpAlaThr***GlnAspProSerSerIle 40

Qy       1141 AAAAATACAAAAAATTA----------------GCCAAGCATGGTGGTGCACGCTTGTAA 1098
Db         41 LysIleLys***Asn***LeuGly********AlaGlyHisGlyGly***HisLeu*** 60

Qy       1097 TTCCAGCTACTCAGGAGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGC 1038
Db         61 SerGlnHisPheGlyMetLeuLys***GluAspCysLeuArgProGly------------ 76

Qy       1037 AGTGAGCTGAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAG 978
Db         77 ------ValGlnAspGlnLeuGlyGlnHisSerLysThrProAlaLeu***Lys***Asn 94

Qy        977 AAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAA 918
Db         95 LysIleSer--------------------------------TrpVal*********--- 102

Qy        917 AGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAA 858
Db        103 ------------GlnGlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCys*** 118

Qy        857 GGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCC 798
Db        119 SerArgArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaArgPro 138

Qy        797 AGTCTCTACAAA---AAAATT--------------------------------TTTTTA 774
Db        139 GlnLeuTyrLysAsnLysIleLysLeuAlaGly*********HisProAlaAsnPheIle 158

Qy        773 ATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCT------------ACTTGGGAG 726
Db        159 LeuPheLeu***SerTrpGlyLeuAlaMetLeuProLysLeuValLeuAsnSerTrpSer 178

Qy        725 GCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGCA 666
Db        179 GlnAlaIle----------------------LeuLeuLeuGlnHisProLysValLeu 190

Qy        665 CCACTG 660
Db        191 GlyLeu 192

RESULT 13
B40201
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC9
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of thi
.

Alignment Scores:
Pred. No.:              1.69e-16      Length:         301
Score:                  239.50        Matches:        81
Percent Similarity:     48.6%         Conservative:   23
Best Local Similarity:  37.9%         Mismatches:     67
Query Match:            5.4%          Indels:         44
DB:                     4             Gaps:           10

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x B40201 (1-301)

Qy        667 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 726
Db        113 SerThrLeuGlyCys***SerArgArgIleAla***AspGlnGluPheLysThrSerLeu 132

Qy        727 TCCCAAGTAGCTGGG----------------------------ACTACGAGCGTGCAC 756
Db        133 GlyAsnIleAlaArgProGlnLeuTyrLysAsnLysIleLysLeuAlaGly********* 152

Qy        757 CACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCC 816
Db        153 HisProAlaAsnPheIle------LeuPheLeu***SerTrpGlyLeuAlaMetLeuPro 170

Qy        817 AGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTG 876
Db        171 LysLeuValLeuAsnSerTrpSerGlnAlaIleLeuLeuLeuGlnHisProLysValLeu 190

Qy        877 GGATTACAGGGGTGAGCCACCATG---TGCGGCTACTTATTTCTTTACATTCCATCTTTC 933
Db        191 GlyLeuGlnMetLeuAlaThrMetProCys---------------------------- 200

Qy        934 CAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTG 993
Db        201 --------*********ThrGln---------LeuIleLeuPheTyr------PheTyr 212

Qy        994 AGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCC 1053
Db        213 ArgAlaGlyValLeuLeuCys---CysProSerTrpSer***------------ThrPro 227

Qy       1054 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 1113
Db        228 GlyLeuLysGlnSerSerTyrPheSerIleProLysCysTrpAspTyrArgCys***Pro 247

Qy       1114 CCATGCTTGGCTAA------------------TTTTTTGTATTTTTAGCAGAGATGGGG 1154
Db        248 ProCysProAla*********ProSer***PheTyrPheIlePheIle---GluLeuGly 266

Qy       1155 TTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCA 1214
Db        267 SerCysTyrValAlaGlnAlaGlyLeuGluLeuLeuValSerSerAsnProProThrSer 286

Qy       1215 GTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCC 1256
Db        287 AlaSerGlnSerAlaGlyIleThrAspValSerHisHisAla 300

RESULT 14
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
```

```
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              2.03e-16      Length:         579
Score:                  239.00        Matches:        125
Percent Similarity:     34.7%         Conservative:   24
Best Local Similarity:  29.1%         Mismatches:     121
Query Match:            5.4%          Indels:         160
DB:                     4             Gaps:           14

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x D40201 (1-579)

Qy         11 TGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAAATGCTTTCAATAA 70
              |||||||||   |||   ||||||
Db        299 CysCysProGlyTrpSerThrGlu---------------------------------- 306

Qy         71 ATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCCTGAAGCATCT 130
                                                                    |||
Db        307 ----------------------------------------------------ValGlnSer 309

Qy        131 TGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGACCTATTGCTCT 190
              |||   :::   |||   |||
Db        310 TrpLeuThrGluPro***Leu------------------------------------- 316

Qy        191 GCACTTACTCCTATCTCAGCTGTCCCT------CCCACTTTCCAGGTGCTGCCAGACACA 244
                   |||   |||:::   |||      |||   |||   |||       |||
Db        317 ---------ProArgSerSerAspProProAlaProAlaPhe***Val-AlaGlyThrTh 333

Qy        245 TGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACAMMCC 304
                  ||||||
Db        333 rGlyLeuCysTyrGln-----------------------------------------Il 339

Qy        305 GTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTAGGTT 364
               :::            |||:::   |||   |||         ||||||
Db        339 eTrpLeuIlePheValSerPhePheCysArgGlyGlyValSerThrCysCys-------- 356

Qy        365 TATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAG 424
                                                    ||||||   :::   ::
Db        357 ---------------------------------------ProGlyCysSer***Th 362

Qy        425 TAACAAC---AGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAA 481
              :         |||          :::          |||||||||      :::
Db        362 rProGluLeuGluHisThrProAlaMetAlaProLysValLeuGlyLeuGlnAla----- 380

Qy        482 TCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACTG 541

Db        380 ---------------------------------------------------------- 380

Qy        542 AGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTG 601
                                  |||      ||||||:::                  ||
Db        381 ----------------------Cys------SerHisAla-----------------Tr 385

Qy        602 GGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACA 661
              |                        :::   ||||||      |||   ||||||      ||
Db        385 p*********---------PheGlnAlaGlySerTrpTyrValAlaGlnAlaGlyValGl 402

Qy        662 GTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCT 721
              |   |||||||||   ||||||  ||||||:::                  ||||||  |||
Db        402 nArgCysAsnHisGlySerLeu-SerLeuAspPheProGlyGlnValIleLeuLeu-Pro 421

Qy        722 CAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGC---TAATTAAAA 778
              ||||||   :::   |||||||||:::|||      |||      |||      |||
Db        422 GlnProPheGlu***LeuGlyLeuGlnAlaTyrAlaThrArgSerGly***PheLeuTyr 441

Qy        779 AAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGC 838
                  ||||||||||||   |||      :::|||   ||||||   |||:::||||||::
Db        442 Leu-PhePheValGluValGlyPheArgHisValAlaGlnAlaValLeuGluLeuLeuSe 461

Qy        839 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA 898
              :   ||| |||   ||||||||| :::   |||||||||||||||||     |||   |
Db        461 rTrpSer-IleHisLeuProTrp-LeuLeuLysCysTrpAspTyrArgArg--AlaAlaM 480

Qy        899 TGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAAC 958
              ||   |||
Db        480 etLeuGly---------------------------------------------------- 482

Qy        959 AGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTC 1018
                                                  ||||||   |||        |||:
Db        483 ------------------------*********PheArgGlnGlyLeuGlyMetLeuProA 495

Qy       1019 AACCTCCGTTCAGCTCACTG-----------CAACCTCTGCCTCCCGGGTTCAAGYGAT 1066
              ::      :::                          |||          ||||||
Db        495 rgLeuGluTyrArgGlyAlaIleMetAlaHis***AlaLeuThrSerGlnValLys***S 515

Qy       1067 TCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTA 1126
                ||||||   |||||||||||||||||:::|||:::       |||   |||||||
Db        515 erSerCysProSerLeuLeuSerSerTrpAspTyrArgLeuMetLeuProAspLeuAlaA 535

Qy       1127 ATTTT---TTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAA 1183
              |||||   ::::::|||      |||||||||||   |||||||||||||      |
Db        535 snPheCysIlePhePheLeu***ArgTrpGlyPheAspMetLeuProArgLeuPheLeuA 555

Qy       1184 ACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGT 1243
              |||||     :::                      |||   ||||||||||   |||||
Db        555 snSer--***AlaGlyAlaTyrThrCysHisGlySer***SerAlaGlyIleThrGlyVa 574

Qy       1244 GAGTCACTGT 1253
              |     |||
Db        574 lGlnProCys 577

RESULT 15
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked r
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.:              8.51e-16      Length:         100
Score:                  232.50        Matches:        51
Percent Similarity:     75.3%         Conservative:   10
Best Local Similarity:  63.0%         Mismatches:     18
Query Match:            5.3%          Indels:         2
DB:                     2             Gaps:           1

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x A46010 (1-100)

Qy       2022 TTTATTTATTTTTTTGAGATAGAGTCTCGCTGTGTCGCCCAGGCTGGAGTGCAGTGGTGG 1963
              |||   :::|||||||||   ||||||      ||||||:::||||||||||||||||
Db          2 PhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCys 21

Qy       1962 GATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTC 1903
              ||||||||||||||:::       ||| ||||||||||:::|||   ||||||||||
Db         22 AspLeuGlySerLeuLysSerProPro-ProGlySerSerAspSerProAlaSerAlaSe 41

Qy       1902 CCGAGTAGCTGGGACTACAGGCATGCATCACTACACCCAGCTAATTTTTGTATTTTTAGT 1843
              ||||||||||||   ||||||||||||||:::|||||||||||||||||||||||||||
Db         41 rArgValAlaGlyIleThrGlyMetHisHisHisThrGlnLeuIlePheValPheLeuVa 61
```

```
Qy       1842 AGAGATAGGGTTTCAC---CATATTGACCAGGCTGGTCTCGAACTTCTCACCTCAAGTGA 1786
              ||||   |||   |||      :::      :::   |||   :::   |||:::   ::
Db         61 lGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAs 81

Qy       1785 T 1785
              :
Db         81 n 81
```

Search completed: February 8, 2006, 11:19:29
Job time : 106.338 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 10:23:31 ; Search time 115.328 Seconds
(without alignments)
3060.024 Million cell updates/sec

Title: US-10-607-806-1-G7328-T9182_COPY_7000_9500
Perfect score: 4426
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/MARTIN10607806/runat_08022006_092018_27658/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=MARTIN10607806_@CGN_1_1_694_@runat_08022006_092018_27658 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	502.5	11.4	591	1	ALU1_HUMAN	P39188 homo sapien
c	2	498	11.3	591	1	ALU1_HUMAN	P39188 homo sapien
	3	479	10.8	593	1	ALU7_HUMAN	P39194 homo sapien
c	4	463	10.5	593	1	ALU7_HUMAN	P39194 homo sapien
	5	443	10.0	593	1	ALU6_HUMAN	P39193 homo sapien
	6	442	10.0	591	1	ALU8_HUMAN	P39195 homo sapien
c	7	439.5	9.9	591	1	ALU8_HUMAN	P39195 homo sapien
c	8	425.5	9.6	593	1	ALU6_HUMAN	P39193 homo sapien
	9	406.5	9.2	587	1	ALU2_HUMAN	P39189 homo sapien
	10	397.5	9.0	375	2	O60448_HUMAN	O60448 homo sapien
	11	396.5	9.0	239	2	Q6ZWA9_HUMAN	Q6zwa9 homo sapien
	12	386	8.7	179	2	Q6ZSI5_HUMAN	Q6zsi5 homo sapien
	13	384.5	8.7	158	2	Q8NAL9_HUMAN	Q8nal9 homo sapien
c	14	382.5	8.6	585	1	ALU5_HUMAN	P39192 homo sapien
	15	382.5	8.6	587	1	ALU3_HUMAN	P39190 homo sapien
	16	381	8.6	241	2	Q6ZTY3_HUMAN	Q6zty3 homo sapien
c	17	377.5	8.5	587	1	ALU2_HUMAN	P39189 homo sapien
	18	372.5	8.4	585	1	ALU5_HUMAN	P39192 homo sapien
	19	372	8.4	200	2	Q6ZUK3_HUMAN	Q6zuk3 homo sapien
c	20	366.5	8.3	156	2	Q8NBH4_HUMAN	Q8nbh4 homo sapien
	21	365	8.2	603	1	ALU4_HUMAN	P39191 homo sapien
	22	363	8.2	314	2	Q6ZTL7_HUMAN	Q6ztl7 homo sapien
c	23	363	8.2	603	1	ALU4_HUMAN	P39191 homo sapien
c	24	361	8.2	146	2	Q6ZRX6_HUMAN	Q6zrx6 homo sapien
c	25	360.5	8.1	128	2	Q6ZPB2_HUMAN	Q6zpb2 homo sapien
c	26	350.5	7.9	129	2	Q6ZUW8_HUMAN	Q6zuw8 homo sapien
c	27	344.5	7.8	150	2	Q6ZPA0_HUMAN	Q6zpa0 homo sapien
c	28	344	7.8	123	2	Q6ZMQ3_HUMAN	Q6zmq3 homo sapien
c	29	342	7.7	129	2	Q8N9K0_HUMAN	Q8n9k0 homo sapien
	30	339.5	7.7	176	2	Q6ZR92_HUMAN	Q6zr92 homo sapien
c	31	336	7.6	124	2	Q6ZSN0_HUMAN	Q6zsn0 homo sapien
	32	335.5	7.6	170	2	Q8NAI3_HUMAN	Q8nai3 homo sapien
c	33	335.5	7.6	587	1	ALU3_HUMAN	P39190 homo sapien
c	34	334	7.5	164	2	Q6ZUK0_HUMAN	Q6zuk0 homo sapien
c	35	334	7.5	375	2	O60448_HUMAN	O60448 homo sapien
c	36	333	7.5	137	2	Q6ZS97_HUMAN	Q6zs97 homo sapien
c	37	331.5	7.5	127	2	Q6ZVW2_HUMAN	Q6zvw2 homo sapien
	38	329.5	7.4	157	2	Q8N845_HUMAN	Q8n845 homo sapien
c	39	329	7.4	167	2	Q6ZSN3_HUMAN	Q6zsn3 homo sapien
c	40	328.5	7.4	120	2	Q6ZTA3_HUMAN	Q6zta3 homo sapien
c	41	328	7.4	140	2	Q6ZP99_HUMAN	Q6zp99 homo sapien
c	42	327.5	7.4	115	2	Q9N083_MACFA	Q9n083 macaca fasc
c	43	327	7.4	131	2	Q6ZTD3_HUMAN	Q6ztd3 homo sapien
	44	326	7.4	164	2	Q6ZTS0_HUMAN	Q6zts0 homo sapien
	45	326	7.4	334	2	Q6ZR97_HUMAN	Q6zr97 homo sapien

ALIGNMENTS

RESULT 1
ALU1_HUMAN
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:              3.51e-42       Length:         591
Score:                  502.50         Matches:        129
Percent Similarity:     48.6%          Conservative:   15
Best Local Similarity:  43.6%          Mismatches:     65
Query Match:            11.4%          Indels:         87
DB:                     1              Gaps:           4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU1_HUMAN (1-591)

Qy        607 TTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGT 666
              ||||||:::          |||||||||||||||||||||||||    |||   ||||||
Db        298 PhePheLeu---------ArgGlnGlyLeuAlaLeuSerProArgLeuGluCysSerGly 314

Qy        667 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 726
              |||      ||||||||||||:::|||     |||   |||          ||||||
Db        315 AlaIleThrAlaHisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla 334

Qy        727 TCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTT 786
              |||:::|||||||||||||||         |||||||||   ||||||        ||||
Db        335 SerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArgLeuIlePheVal--PheP 354

Qy        787 TTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCA 846
              ||||||||||||||   :::||||||   ||||||   |||:::||||||   ||||||:
Db        354 heValGluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGluLeuLeuGlySerSerA 374

Qy        847 ATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGG- 905
              ::||||||      ||||||||||||||||||||||||||||||||||||      |||
Db        374 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg* 394

Qy        905 ---------------------------------------------------------- 905

Db        394 ********PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAlaValAlaA 414

Qy        906 ----------------------------------------------CTACTTATTTCTTTAC 921
                                                           ||||||
Db        414 rgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnProP 434

Qy        922 ATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTT- 980
                          ::::::   :::           |||         |||||||||
Db        434 roGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu* 454

Qy        980 ---------------------------------------------------------- 980

Db        454 **ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIleL 474

Qy        981 ----------------------------------------------------------C 981

Db        474 euProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly**** 494

Qy        982 CTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC--------- 1032
                    ||||||||||||   |||      :::            |||
Db        494 *****PhePheGluThrGlySerArgSerValAlaGlnAlaGlyVal-GlnTrpArgAsp 513

Qy       1033 ------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTG 1086
                    ||||||||||||   |||||||||   |||      ||||||||||||||||
Db        514 HisGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuPro 533

Qy       1087 AGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAG 1146
              |||||||||:::|||:::||||||||||||      ||||||  ||||||||||||   ||
Db        534 SerSerTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheCysAr 553

Qy       1147 AGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGC 1206
              |||||||||         ||||||||||||||||||:::||||||    ||||||   |||::
Db        553 gAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSe 573

Qy       1207 CTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAG 1246
              :   |||   ||||||||||||||:::||||||||||||
Db        573 rArgLeuGlyLeuProLysCysTrpAspTyrArgArgGlu 586

RESULT 2
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
```

```
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:              1.02e-41        Length:         591
Score:                  498.00          Matches:        125
Percent Similarity:     61.2%           Conservative:   12
Best Local Similarity:  55.8%           Mismatches:     53
Query Match:            11.3%           Indels:         34
DB:                     1               Gaps:           2

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU1_HUMAN (1-591)

Qy       1261 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1202
                 |||:::|||:::   |||||||||:::|||||||||||||||||||   |||||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       1201 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 1142
                 |||||||||   ||||||   |||||||||||||||||||      ||||||:::|||
Db          21 SerLeuGluProArgSerSerArgProAlaTrpAlaThr******AsnProValSerThr 40

Qy       1141 AAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGA 1082
                 ||||||||| |||||||||:::|||||||||        |||:::|||||||||:::||
Db          41 LysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValValProAlaThrArgGl 60

Qy       1081 GGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCT-------- 1030
                 ||||   ||||||   |||||||||||||||||   |||||||||   |||
Db          60 uAlaGluAlaGlyGlySerLeuGluProGlyArgSerArgLeuGln***AlaValIleAl 80

Qy       1029 -------GAACGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 977
                        ||| ||||||      |||   ||||||||||||
Db          80 aProLeuHisSerSerLeu-GlyAspArgAlaArgProCysLeuLysLys--*******-- 98

Qy        976 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 917

Db         98 ------------------------------------------------------------ 98

Qy        916 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 857
                    |||      |||||||||   |||   ||||||||||||||   |||||
Db         99 -------***AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 116

Qy        856 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 797
                |  ||||||   |||   |||||||||   |||:::|||||||||   ||:::|||
Db        116 gArgGluAspHisLeuSerProGlyValArgAspGlnProGlyGlnHisSerGluThrPr 136

Qy        796 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 737
                 |||||||||  |||          ||||||   ||||||||||||||      ||||||
Db        136 oSerLeuGln--LysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGln 155

Qy        736 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 677
                 |||||||||||||||||      |||   ||||||   ||||||   |||||||||:::
Db        156 LeuLeuGlyArgLeuArgGlnGluAspArgLeuSerProGlyGlyArgGlyCysSerGlu 175

Qy        676 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 617
                      ||||||      ||||||   ||||||:::|||||||||||||||:::
Db        176 Pro***SerArgHisCysThrProAlaTrpAlaThrGluArgAspProValSerLysLys 195

RESULT 3
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
```

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:              9.39e-40       Length:         593
Score:                  479.00         Matches:        117
Percent Similarity:     58.5%          Conservative:   7
Best Local Similarity:  55.2%          Mismatches:     57
Query Match:            10.8%          Indels:         31
DB:                     1              Gaps:           2

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU7_HUMAN (1-593)

Qy        642 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||   ||||||      ||||||   :::      ||||||||||||   ||||||
Db        308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProGly 327

Qy        702 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 761
                 |||      :::   ||||||||||||||||||||||||||   |||   ||||||
Db        328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProPro 347

Qy        762 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 821
              |||   ||||||          |||   ||||||   |||         |||||||||
Db        348 ArgProAlaAsn--PheCysIlePheSerArgAspGlyValSerProCysTrpProGlyT 367

Qy        822 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 881
                |||   ||||||      :::   |||:::   ||||||:::||||||||||||||||
Db        367 rpSerArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspT 387

Qy        882 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 941
              |||||   |||||||||      |||         |||
Db        387 yrArgArgGluProProArgProAla*********Phe---------------------- 399

Qy        942 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAG 1001
                                                                |||   |||   |
Db        400 ----------------------------------------------Phe***AspGlyV 404

Qy       1002 TCTCACTTCATCACCTCAACCTCCGT--------------TCAGCTCACTGCAACCTCT 1046
              |||||||           |||                 ||||||||||||||||||
Db        404 alSerLeuLeuLeuProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuA 424

Qy       1047 GCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCG 1106
                |||||||||||||||||||||||||||   ||||||   ||||||||||||||||
Db        424 rgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyA 444

Qy       1107 TGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTT 1166
                |||||||||   ||||||  ||||||||||||   |||   ||||||:::||||||
Db        444 laArgHisHisAlaArgLeuIle-PheValPheLeuValGluThrGlyPheHisHisVal 463

Qy       1167 GCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGT 1226
                ||||||||||:::||||||||||   |||   |||||||||   |||||||||
Db        464 GlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspProProAlaSerAlaSerGlnSer 483

Qy       1227 GCTGGAATTATAGGCGTGAGTCACTGTGCC 1256
              |||||||||   ||||||||||||   |||
Db        484 AlaGlyIleThrGlyValSerHisArgAla 493

RESULT 4
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
```

```
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:              4.22e-38        Length:         593
Score:                  463.00          Matches:        118
Percent Similarity:     58.8%           Conservative:   15
Best Local Similarity:  52.2%           Mismatches:     58
Query Match:            10.5%           Indels:         35
DB:                     1               Gaps:           3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU7_HUMAN (1-593)

Qy       1261 GGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGA 1202
                 |||:::|||:::   |||||||||:::|||||||||||||||||||   |||||||||
Db          1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20

Qy       1201 TCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCT 1142
                 |||   ||||||||||||   ||||||||||||   ||||||   ||||||:::|||
Db         21 SerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThr 40

Qy       1141 AAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGA 1082
                 |||||||| |||||||||:::|||||||||         |||||||||||||||:::||
Db         41 LysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGl 60

Qy       1081 GGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAA----- 1027
                 ||||   ||||||||||||||||||||||||:::|||||||||   ||||||
Db         60 uAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAl 80

Qy       1026 ----------CGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGA 977
                           |||         |||||||||   |||||||||
Db         80 aProLeuHisSerSerLeuGlyAsnLysSerGluThrProSerGlnLys---------- 96

Qy        976 AAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAA 917
                                                                      ||
Db         97 ----------------------------------------------------------Ly 97

Qy        916 GAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAG 857
                 |          |||      |||||||||   |||   |||||||||||||||   |||||
Db         97 s*********AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProAr 117

Qy        856 GTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCA 797
                 |   |||   |||:::   ||||||   |||:::|||||||||      :::|||
Db        117 gArgValAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrPr 137

Qy        796 GTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAG 737
                 ||||||     |||          ||||||   |||||||||   |||      ||||||
Db        137 oSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyGlyArgLeu***SerGln 156

Qy        736 CTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGAC 677
                 ||||||||||||||||||      :::   |||:::   ||||||   |||||||||:::
Db        157 LeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGlu 176

Qy        676 CTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 617
                      ||||||      ||||||   |||              :::|||
Db        177 ProArgSerArgHisCysThrProAlaTrpAlaThrArg-----------AlaLysLeu 192

Qy        616 CAACAAAAAAAA 605
                 :::   ||||||
Db        193 ArgLeuLysLys 196

RESULT 5
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
```

```
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              4.91e-36      Length:         593
Score:                  443.00        Matches:        113
Percent Similarity:     56.6%         Conservative:   7
Best Local Similarity:  53.3%         Mismatches:     61
Query Match:            10.0%         Indels:         31
DB:                     1             Gaps:           2

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU6_HUMAN (1-593)

Qy        642 GTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAG 701
              |||   ||||||      ||||||   :::      |||   ||||||   ||||||
Db        308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProProGly 327

Qy        702 CTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCA 761
                 |||      :::   |||||||||||||||||||||||||||      |||||||||
Db        328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProPro 347

Qy        762 CGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCT 821
              |||   ||||||           |||   ||||||   |||         |||   |||
Db        348 ArgProAlaAsn--PheCysIlePheSerArgAspGlyValSerProCysTrpSerGlyT 367

Qy        822 TGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATT 881
                |||   ||||||      :::   |||:::   ||||||:::|||||||||||||||
Db        367 rpSerArgThrProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrpAspT 387

Qy        882 ACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAA 941
              |||||   |||||||||      |||          |||
Db        387 yrArgArgGluProProArgProAla*********Phe---------------------- 399

Qy        942 TGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAG 1001
                                                              |||   |||   |
Db        400 ----------------------------------------------Phe***AspGlyV 404

Qy       1002 TCTCACTTCATCACCTCAACCTCCGT--------------TCAGCTCACTGCAACCTCT 1046
```

```
              ||||||||          |||                   |||||||||   ||||||
Db        404 alSerLeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuA 424

Qy       1047 GCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCG 1106
              ||||||||||||||||||||||||||   ||||||   ||||||||||||||||   :
Db        424 rgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyM 444

Qy       1107 TGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTT 1166
              ::   |||||||||   ||||||||||||||   |||   ||||||   ||||||
Db        444 etArgHisHisAlaArgLeuIle-PheValPheLeuValGluThrGlyPheLeuHisVal 463

Qy       1167 GCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGT 1226
                 ||||||||||||:::|||   ||||||   |||   |||||||||   |||||||||
Db        464 GlyGlnAlaGlyLeuGluLeuProThrSerGlyAspProProAlaSerAlaSerGlnSer 483

Qy       1227 GCTGGAATTATAGGCGTGAGTCACTGTGCC 1256
              |||||||||   ||||||||||||   |||
Db        484 AlaGlyIleThrGlyValSerHisArgAla 493

RESULT 6
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
```

```
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:                6.22e-36        Length:          591
Score:                    442.00          Matches:         172
Percent Similarity:       43.2%           Conservative:    37
Best Local Similarity:    35.5%           Mismatches:      145
Query Match:              10.0%           Indels:          130
DB:                       1               Gaps:            17

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU8_HUMAN (1-591)

Qy          2 TCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAAC---TCTGTCTCAAAAAAA----- 53
              |||   ||||||   |||||||||   ||||||       ||||||||||||||||
Db        178 SerArgHisCysThrProAlaTrpAlaThrGlu-ArgAspSerValSerLysLys***** 197

Qy         54 ----------------------------------------AAAAATGCTTTCAATAAATA 73
                                                      |||   :::
Db        197 ****ArgAlaGluAlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgLy 217

Qy         74 TATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCCTGAAGCATCTT-- 131
               :::      |||      |||      ::::::   :::   :::          |||
Db        217 sIleThr***GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProArgLeuTy 237

Qy        132 ----------------------GGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCA-- 167
                                     |||   |||         |||:::      ::::::
Db        237 r***LysTyrLysAsn***ProGlyValValAlaArgAlaCysAsnProSerTyrSerGl 257

Qy        168 -GGTGAGGGCTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTC 226
               |||   |||      ||||||                 :::      :::   :::||
Db        257 yGly***GlyArgArgIleAla---------***ThrArgGluAlaGluValAlaValSe 274

Qy        227 CAGGTGCTGCCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAA 286
               ||||                          |||:::|||      |||         ::
Db        274 rArg-------------------------AspArgAlaThrAlaLeuGlnProGlyAr 285

Qy        287 A-----------------------------------TTTCTGCTGGACAMMCCGTACAC 310
              :                                   |||   |||
Db        285 gGlnSerGluThrProSerGlnLysLys*********PhePheLeuArgArgSerLeuAl 305

Qy        311 CCACACCTATTCATACTCGTGCTCTGGCTCG---GCAATCACCTGTAGCAGTAGGTTTAT 367
                  :::          |||||||||:::   :::      |||:::   |||
Db        305 aLeuSerProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuArgLeu-- 324

Qy        368 CCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAA 427
                                                 ||||||   ::::::|||
Db        325 -----------------------------------------ProGlySerSerAspSerPr 331

Qy        428 CAACAGC---------------------------------CATGATTTAGTGTTAATTTT 454
                  |||                                  |||         ||||||||
Db        331 oAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisHisAlaArgLeuIlePh 351

Qy        455 CTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTACAATT 514
              |:::   |||            |||            :::   :::|||         ||
Db        351 eValPheLeuValGluThrGlyPhe-----------HisHisValGlyGlnAlaGlyLe 367

Qy        515 ATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATC 574
              |     |||                     :::   ||| :::|||    ||||||:::
Db        367 uGluLeuLeu-------------------ThrSerGly-AspLeu-ProProLysVal 379

Qy        575 ACACAGCTT---------GTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGA 625
                  |||           :::         |||                    |||
Db        380 LeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro*********PhePhe*** 399

Qy        626 GACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCA 685
              |||   |||   ||||||   ||||||   |||   |||   :::       :::||||||
Db        400 AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArgLeuThrAla 419

Qy        686 GCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTA 745
                 |||   |||      |||   |||      |||||||||||:::   |||||||||
Db        420 ThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProGlu***LeuGlyLeu 439

Qy        746 CGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTA 805
              :::|||      ||||||||||||          |||       |||   |||   |
Db        440 GlnAlaArgAlaThrThrProGly***PheLeuTyrPhe******--ArgArgGlyPheT 459

Qy        806 CTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATC 865
              ||   |||||||||||||||   ||||||       |||   |||   |||        |
Db        459 hrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeu---------P 476

Qy        866 CCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC 925
              || ||||||||||||||||   |||||||||      |||:::
Db        476 ro-LysCysTrpAspTyrArgArgGluProProArgProAlaSer-AlaArg******** 495

Qy        926 CATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTT 985

Db        495 *---------------------------------------------------------- 495

Qy        986 CTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC------------- 1032
              ||||||||||||||||||      :::      |||
Db        496 -PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArgAspLeuG 515

Qy       1033 --TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTA 1090
                |||||||||||   ||||||||||||||||   ||||||||||||||||||   ||||
Db        515 lySerLeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerS 535

Qy       1091 GCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGAT 1150
                |||||:::|||:::|||||||||||||      ||||||  ||||||||||||||||||||
Db        535 erTrpAspTyrArgArgAlaProProArgProAlaAsn-PheCysIlePheSerArgAsp 554

Qy       1151 GGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGC 1210
                ||||||   ||||||   ||||||||||||:::||||||||||||:::   |||
Db        555 GlyValSerProCysTrpProGlyTrpSerArgThrProAspLeuArg***Ser------ 572

Qy       1211 CTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
                 ||| ||||||||||||||||||   ||||||||||||   |||   |||
Db        573 ---Ser-SerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgPro 588

RESULT 7
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;
```

```
Alignment Scores:
Pred. No.:              1.13e-35      Length:          591
Score:                  439.50        Matches:         166
Percent Similarity:     38.5%         Conservative:    30
Best Local Similarity:  32.6%         Mismatches:      131
Query Match:            9.9%          Indels:          182
DB:                     1             Gaps:            15

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU8_HUMAN (1-591)

Qy      1262 CGGCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAG 1203
             |||||||||   |||   ||||||   |||   |||||||||         ||| |||||
Db         4 ArgProGlyAlaValAlaHisAlaCysAsnProSerThrLeu---------Gly-GlyAr 20

Qy      1202 ATCACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGC 1143
             ||||   ||||||||||||   ||||||||||||   ||||||   |||||||:::|||
Db        20 gSerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerTh 40

Qy      1142 TAAAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGG 1083
              ||||||||| |||||||||:::|||||||||         |||||||||||||||:::|
Db        40 rLysAsnThr-LysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgG 60

Qy      1082 AGGCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAA---- 1027
             |||||   |||||||||||||||||||||||:::|||||||||   ||||||
Db        60 luAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleA 80

Qy      1026 -----------CGGAGGTTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAG 978
                        ||| ||||||      ||||||   |||||||||
Db        80 laProLeuHisSerSerLeu-GlyAspArgAlaArgLeuArgLeuLysLys********* 99

Qy       977 AAAATAGGCAGTAATCCCTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAA 918

Db        99 ------------------------------------------------------------ 99

Qy       917 AGAAATAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAA 858
                     |||      ||||||||||||||||||||||||||||||
Db       100 AlaGly-GlyGlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrp-Glu---- 117

Qy       857 GGTAGGAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCC 798
                   ||||||   |||:::   ||||||   |||:::|||||||       :::|||
Db       118 -----GluAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrP 136

Qy       797 AGTCTCTACAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCA 738
               ||||||   |||          ||||||   ||||||||||||||       |||||
Db       136 roSerLeu--LeuLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGl 155

Qy       737 GCTACTTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGA 678
             ||||||||||||||||||      :::   |||:::   ||||||   |||||||::
Db       155 nLeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGl 175

Qy       677 CCTATGGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACA 618
             :            ||||||   ||||||:::|||   |||||||::
Db       175 uProArgSerArgHisCysThrProAlaTrpAlaThrGluArgAspSerValSerLysLy 195

Qy       617 A------------------------------------------------------------ 617
             :
Db       195 s*********ArgAlaGluAlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGl 215

Qy       616 ------------------------------------------CAACAAAAAAAAACCCA 600
                                                       |||   ::
Db       215 yArgLysIleThr***GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProAr 235

Qy       599 AACTGTCACCACTTACAAGCTGTGTGATCTTGGGCAAATAGCTCAGCCTTCTGAGCCTCA 540
             :            ||||||         |||   :::|||:::         |||
Db       235 gLeuTyr***LysTyrLysAsn***Pro---GlyValValAlaArgAlaCysAsnProSe 254

Qy       539 GTTTTCT-------------------------------------------TATCTGTTAAG 522
              :::|||                                            :::||
Db       254 rTyrSerGlyGly***GlyArgArgIleAla***ThrArgGluAlaGluValAlaValSe 274

Qy       521 TGAGGATAATTGTACCTATCCCATAGTGTTGTTCTGAGGATTAAAGGAGACACTGCCCAG 462
             |   |||      |||           |||      |||   :::         |||
```

```
Db        274 rArgAspArgAlaThr---------AlaLeuGlnProGlyArgGlnSerGluThrProSe 291

Qy        461 AACCAAGAAAATTAACACTAAATCATGGCTGTTGTTACTATTATTTCCCCCCGGCCTACT 402
                  ||||||
Db        291 rGlnLysLys------------------------------------------------- 294

Qy        401 GAGAACCAACTAGAATTCATAGGTCAAGGAAGGGATAAACCTACTGCTACAGGTGATTGC 342

Db        294 ------------------------------------------------------------ 294

Qy        341 CGAGCCAGAGCACGAGTATGAATAGGTGTGGGTGTACGGKKTGTCCAGCAGAAATTTACA 282

Db        295 ------------------------------------------------*********PhePh 299

Qy        281 GCTGTCCAGCTTCTTGGCCTGGTCRTAGCAGTTGTCATGTGTCTGGCAGCACCTGGAAAG 222
                |||         ||||||   |||   :::|||   |||                 ||
Db        299 eLeuArgArgSerLeuAlaLeuSerProArgLeuGluCys-----------------Se 313

Qy        221 TGGGAGGGACAGCTGAGATAGGAGTAAGTGC------AGAGCAATAGGT---CAGCCCTC 171
                ||||     |||         :::   |||     |||     |||         ||
Db        313 rGlyAlaIleSer---------AlaHisCysAsnLeuArgLeuProGlySerSerAspSe 330

Qy        170 ACCTGCCCACTCTCAGGAACAGGTGGGGATGACTTCGCCAAGATGCTTCAGGAGA----- 116
                |||||||   :::             |||:::|||     |||         |||
Db        330 rProAlaSerAlaSerArgValAlaGlyIleThrGlyAlaArgHisHisAlaArgLeuIl 350

Qy        115 ------------------------------------------------AATGATCCTAT 105
                                                               :::   ::
Db        350 ePheValPheLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGluLeuLe 370

Qy        104 GGCTTGAAAAAATATAAGTCCTTTTATCATATATTTATTG---------------------- 65
                :         ::::::   |||   ::::::   |||
Db        370 uThrSerGlyAspLeuProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGl 390

Qy         64 ----------AAAGCATTTTTTTTTTTTTGAGACAGAGTTGCTCTGTTACCCAGGCTGGAC 15
                           ||||||   :::         ||||||   |||||||||::
Db        390 yLeuArgPro*********PhePhe***AspGlyValSerLeuCysArgProGlyTrpSe 410

Qy         14 AGCAGTGACACGA 2
              :||||||   |||
Db        410 rAlaValAlaArg 414
```

```
RESULT 8
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              3.15e-34       Length:         593
Score:                  425.50         Matches:        164
Percent Similarity:     40.0%          Conservative:   36
Best Local Similarity:  32.8%          Mismatches:     138
Query Match:            9.6%           Indels:         162
DB:                     1              Gaps:           18

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU6_HUMAN (1-593)

Qy       1260 GCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGAT 1201
                 |||             |||||||||      ||||||||||||||   |||:::||||||
Db        101 AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAsp 120

Qy       1200 CACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTA 1141
                 ||||||||||||||||   |||||||||   |||||||||:::|||||||||||||||
Db        121 HisLeuArgSerGlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeu 140
```

```
Qy       1140 AAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAG 1081
Db        141 LysIleGln-LysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGlnLeuLeuGlyAr 160

Qy       1080 GCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGG 1021
Db        160 gLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysGlyGlu----------- 176

Qy       1020 TTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCC 961
Db        177 ----------------------------------------------------------Pr 177

Qy        960 CTGTTCTGTGGATCTTACATTCTATTGG-----------AAAGATGGAATGTAAAGAAA 913
Db        177 oArgSerArgHisCysThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys** 197

Qy        912 TAAGTAGCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAG 853
Db        197 *******ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGl 217

Qy        852 GAGGATTGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCT 793
Db        217 yArgIleThr***GlyArgGluPheGluThrSerLeuThrAsnMetGluLysProArgLe 237

Qy        792 CTACAAAAAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTAC 733
Db        237 uTyr***Lys--TyrLysAsn***ProGlyValValAlaHisAlaCysAsnProSerTyr 256

Qy        732 TTGGGAGGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTAT 673
Db        257 SerGlyGly***GlyArgArgIleAla***ThrArgGluAlaGluValAlaValSerArg 276

Qy        672 GGTTGCACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAACAAC 613
Db        277 AspArgAlaIleAlaLeuGlnProGlyGlnGlnGluArg-----------AsnSerVal 292

Qy        612 AAAAAAAAA---------------------------------------------------- 604
Db        293 SerLysLys*********PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGly 312

Qy        603 ------------------------CCCAAACTGTCACCA-----------CTTACAAGC 580
Db        313 ValGlnTrpArgAspLeuGlySerProGlnProProProProGlyPheLysArgPheSer 332

Qy        579 TGTGTGATCTTGGGCAAATAGCTCAGCCTTCTGAGC----------------------- 544
Db        333 Cys-----------------LeuSerLeuProSerSerTrpAspTyrArgHisAlaPro 346

Qy        543 ---------CTCAGTTTTCTTATCTGTTAAGTGAGGATAATTGTACCT----------- 505
Db        347 ProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpSerGly 366

Qy        504 ----------------------------------------ATCCCATAGTGTTGTTCT 487
Db        367 TrpSerArgThrProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrp--- 385

Qy        486 GAGGATTAAAGGAGACACTGCCCAGAACCAAGAAAATTAACACTAAATCATGGCTGTTGT 427
Db        386 ---AspTyrArgArgGluProProArgProAla*********Phe-Phe***AspGlyVa 404

Qy        426 TACTATTATTTCCCCCCGG--------------------CCTACTGAGAACCAACTAGA 388
Db        404 lSerLeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuAr 424

Qy        387 ATTCATAGGTCAAGGAAGGGATAAACCTACTGCT--------------------ACAGG 349
Db        424 gLeuProGlySerSer---AspSerProAlaSerAlaSerArgValAlaGlyIleThrGl 443

Qy        348 TGATTGCCGAGCCAGAGCACGAGTATGAATAGGTGTGGGTGTACGGKKTGTCCAGCAGAA 289
Db        443 y---MetArgHisHisAlaArgLeu---------IlePheValPheLeuValGluThrGl 459

Qy        288 ATTTACAGCTGTCCAGCTTCTTGGCCTGGTCRTAGCAGTTGTCATGTGTCTGGCAGCACC 229
```

```
Db        459 yPheLeuHisValGlyGlnAlaGlyLeuGluLeuProThr-Ser---------------- 473

Qy        228 TGGAAAGTGGGAGGGACAGCTGAGATAGGAGTAAGTGCAGAGCAATAGGTCAGCCCTCAC 169
Db        474 ----------------------------------------------GlyAspPro---P 477

Qy        168 CTGCCCACTCTCAGGAACAGGTGGGGATGACTTCGCCAAGATGCTTCAGGAGAAATGATC 109
Db        477 roAlaSerAlaSerGlnSerAlaGlyIleThrGly------------------------- 488

Qy        108 CTATGGCTTGAAAAAATATAAGTCCTTTTATCATATATTTATTGAAAGCATTTTTTTTTT 49
Db        489 ----------------------------ValSerHisArgAlaArg*********PheP 499

Qy         48 TTGAGACAGAGTTG---CTCTGTTACCCAGGCTGGACAGCAGTGACACGA 2
Db        499 heGluThrGluPheArgSerCysCysProGlyTrpSerAlaMetAlaArg 515

RESULT 9
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
```

```
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:              2.88e-32      Length:        587
Score:                  406.50        Matches:       126
Percent Similarity:     47.2%         Conservative:  18
Best Local Similarity:  41.3%         Mismatches:    67
Query Match:            9.2%          Indels:        94
DB:                     1             Gaps:          7

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU2_HUMAN (1-587)

Qy        607 TTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGT 666
              ||||||:::         |||:::   ||||||||||||||||||   |||   ||||||
Db        296 PhePheLeu---------ArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGly 312

Qy        667 GCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCC 726
              |||        ||||||   |||   |||      ||||||          ||||||
Db        313 AlaIleSerAlaHisCysLysLeuArgLeuProGlySerArgHisSerProAlaSerAla 332

Qy        727 TCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTT 786
              |||:::|||||||||||||||         |||||||||   ||||||         |||
Db        333 SerArgValAlaGlyThrThrGlyAlaArgHisHisAlaArgLeuIlePheVal--PheL 352

Qy        787 TTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCA 846
              ||||||||||||   :::   |||   |||      |||   ||||||   |||
Db        352 euValGluThrGlyPheHisArgValSerGlnAspGlyLeuAspLeuLeuThrSer***- 371

Qy        847 ATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA---TGTGC 903
              |||:::   ||||||:::||||||||||||||||||||||   |||||||||      ||
Db        372 -SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgProAl 391

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG--------------A 948
              |      ||||||         :::||| :::   |||:::              :
Db        391 a*********PhePhe***AspGlyValSer-Leu---CysArgProGlyTrpSerAlaV 410

Qy        949 TCCACAGAACAGGGATTACTGCC---------------------------------- 971
              ::   |||:::   :::||||||                                  .
Db        410 alAlaArgSerArgLeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProG 430

Qy        972 ----------------------------------------------TATTTTCTTCCTT 984
                                                           ||||||   |
Db        430 lnProProGlu***LeuGlyLeuGlnAlaProAlaThrThrProGly***PheLeuTyrP 450

Qy        985 TC-------------------------------------------------------- 986
              ||
Db        450 he******ArgArgGlyPheThrValLeuAlaArgMetValSerIleSer***ProArgA 470

Qy        986 ---------------------------------------------------------- 986

Db        470 spProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg* 490

Qy        987 --------TTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC------ 1032
                      ||||||||||||||||||      :::         |||
Db        490 ********PhePheGluThrGluSerArgSerValAlaGlnAlaGlyVal-GlnTrpArg 509

Qy       1033 ---------TCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 1083
                       |||||||||      ||||||||||||      ||||||||||||||||||
Db        510 AspLeuGlySerLeuGlnAlaProProProGlyPheThrProPheSerCysLeuSerLeu 529

Qy       1084 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAG 1143
                |||||||||:::|||:::|||   ||||||      ||||||  ||||||||||||||
Db        530 ProSerSerTrpAspTyrArgArgProProProArgProAlaAsn-PheCysIlePheSe 549

Qy       1144 CAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATC 1203
              ||||||||||||   ||||||   ||||||||||||::::::|||||||||  ||||||
Db        549 rArgAspGlyValSerProCys***ProGlyTrpSerArgSerProAspLeu--ValIle 568

Qy       1204 TGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
                 |||:::   ||||||||||||   |||   |||      |||   ||||||
Db        569 ArgProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 587

RESULT 10
O60448_HUMAN
ID   O60448_HUMAN PRELIMINARY;      PRT;   375 AA.
AC   O60448;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Neuronal thread protein AD7c-NTP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RX   MEDLINE=98064067; PubMed=9399956;
RA   Monte S.M., Ghanbari K., Frey W.H., Beheshti I., Averback P.,
RA   Hauser S.L., Ghanbari H.A., Wands J.R.;
RT   "Characterization of the AD7C-NTP cDNA expression in Alzheimer's
RT   disease and measurement of a 41-kD protein in cerebrospinal fluid.";
RL   J. Clin. Invest. 100:3093-3104(1997).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Neuronal;
RA   de la Monte S.M., Wands J.R.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF010144; AAC08737.1; -; mRNA.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0006915; P:apoptosis; TAS.
DR   GO; GO:0007417; P:central nervous system development; TAS.
SQ   SEQUENCE   375 AA;  41720 MW;  955443950A5BFEFD CRC64;

Alignment Scores:
Pred. No.:              2.14e-31      Length:        375
Score:                  397.50        Matches:       105
Percent Similarity:     56.0%         Conservative:  12
Best Local Similarity:  50.2%         Mismatches:    60
Query Match:            9.0%          Indels:        32
DB:                     2             Gaps:          4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x O60448_HUMAN (1-375)

Qy        597 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGA 656
              :::::::::|||           |||               ||||||||||||||||
```

```
Db        194 IlePheIlePheIlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGly 213

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716

Db        214 ValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSer 233

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776

Db        234 CysProSerLeuLeuSerSerTrpAspTyrArgArg---ProProArgLeuAlaAsnPhe 252

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836

Db        253 PheValPheLeuValGlu-MetGlyPheThrMetPheAlaArgLeuIleLeuIleSer-- 271

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896

Db        272 GlyProCysAspLeuProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 291

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956

Db        292 HisAlaArgLeuIlePheAsnPhe--------------------Cys------------ 300

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016

Db        301 ----------------------------LeuPheGluMetGluSerHisSerValThr 310

Qy       1017 TCAA--------------CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 1061

Db        311 -GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLy 330

Qy       1062 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 1121

Db        330 sArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeuProProHisPr 350

Qy       1122 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 1181

Db        350 oAlaAsn-PheCysIlePheIleArgGlyGlyValSerProTyrLeuSerGlyTrpSerG 370

Qy       1182 AAACTCCTGACCTCAAG 1198

Db        370 lnThrProAspLeuArg 375

RESULT 11
Q6ZWA9_HUMAN
ID   Q6ZWA9_HUMAN PRELIMINARY;      PRT;   239 AA.
AC   Q6ZWA9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ41367.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Caudate nucleus;
RA   Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA   Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK123361; BAC85595.1; -; mRNA.
SQ   SEQUENCE   239 AA;  25635 MW;  409CED42F0603D85 CRC64;

Alignment Scores:
Pred. No.:              2.37e-31       Length:        239
Score:                  396.50         Matches:       110
Percent Similarity:     56.0%          Conservative:  25
Best Local Similarity:  45.6%          Mismatches:    68
Query Match:            9.0%           Indels:        38
DB:                     2              Gaps:          4

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q6ZWA9_HUMAN (1-239)

Qy        625 AGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGC 684

Db          2 ArgTrpSerLeuThrLeuSerProArgLeuGluCysSerGlyAlaIleSerThrAspSer 21

Qy        685 AGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACT 744

Db         22 SerLeuHisLeuProGlySerSerAsnSerLeuThrSerAlaSerGlnValGlyGlyIle 41

Qy        745 ACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTT 804

Db         42 ThrGlyAlaCysHisHisAlaArgLeuIlePheVal--PheSerValGluThrGlyPheH 61

Qy        805 ACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCAT 864

Db         61 isHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerAsnAspLeuProThrSerAlaS 81

Qy        865 CCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATT 924

Db         81 erGlnAsnAlaGlyIleThrGlyValSerHisArgAlaGln------------------- 94

Qy        925 CCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTT 984

Db         95 --------ProGluLeuCys-----------------------------------------P 99

Qy        985 TCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA------------------------ 1019

Db         99 heIleLeuLysSerPheSerPhePheLeuArgGlnGlyLeuCysHisProGlyTrpSerA 119

Qy       1020 -----ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGC 1074

Db        119 laValAlaGlnSerGlnLeuThrGluAlaSerValSerGlnAlaGlnAlaIleLeuLeuP 139

Qy       1075 CTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTG 1134

Db        139 roGlnProProGlu--TyrGlySerThrGlyMetTyrHisHisAlaTrpLeuIlePheTr 158

Qy       1135 TATTT---TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGA 1191

Db        158 p-GlyGlyLeuValAsnLysGlySerHisTyrIleAlaGlnAlaGlyLeuGluLeuLeuA 178

Qy       1192 CCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACT 1251

Db        178 snSerSerAsnProProAlaSerValSerGlnLysAlaGlyIleIleAspIleSerHisC 198

Qy       1252 GTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGA 1304

Db        198 ysAlaGlnProIleLeuGluSerLeuMetPheIleProAsnProLysGlyArg 215

RESULT 12
Q6ZSI5_HUMAN
ID   Q6ZSI5_HUMAN PRELIMINARY;      PRT;   179 AA.
AC   Q6ZSI5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ45499.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Thalamus;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
```

```
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK127408; BAC86963.1; -; mRNA.
SQ   SEQUENCE   179 AA;  19118 MW;  52286CCCA5A4458E CRC64;

Alignment Scores:
Pred. No.:              2.63e-30        Length:         179
Score:                  386.00          Matches:        109
Percent Similarity:     60.2%           Conservative:   15
Best Local Similarity:  52.9%           Mismatches:     50
Query Match:            8.7%            Indels:         33
DB:                     2               Gaps:           6

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q6ZSI5_HUMAN (1-179)

Qy        633 TCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAA 692
              |||   ||||||:::||||||      ||||||   :::      ||||||||||||
Db          3 SerSerSerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProPro 22

Qy        693 CCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGT 752
              |||         |||      :::   |||||||||   ||||||||||||||||
Db         23 ProSerGlyPheLysArgPheSerCysLeuSerLeuGlnSerSerTrpAspTyrArgHis 42

Qy        753 GCACCACCACGCCTGGCTAATTAAAAAAATTTTTTT---------GTAGAGACTGGGTCT 803
              |||||||||      |||                |||          ||||||||||||
Db         43 AlaProProCysProAlaGly------------PheCysLeuCysValGluThrGlyPhe 58

Qy        804 TACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCA 863
              ::::::||||||||||||   ||||||||||||   ||||||:::||||||   ||||||
Db         59 HisHisValGlyGlnAlaSerLeuLysLeuLeuThrSerSerAspProProAlaLeuAla 78

Qy        864 TCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACAT 923
              ||||||   ||||||||||||||:::   ||||||   :::   |||:::   |||
Db         79 SerGlnIleAlaGlyIleThrGlyMetIleHisHisAlaGlnProLeuLeuPheLeu--- 97

Qy        924 TCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCT 983
                                        |||   |||:::|||         ||||||
Db         98 -----------------------------ArgGlnGlyValThr------LeuLeuPro 105

Qy        984 TTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA-GCTCACTGCAAC 1042
                                          |||::::::::::   |||||||||
Db        106 ---------------------ArgLeuGluCysSerSerAlaIleLeuAlaHisCysIle 118

Qy       1043 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 1102
              ||||||||||||||||||||||||||||||||   ||||||   |||||||||   |||
Db        119 LeuCysLeuProGlySerSerAspSerProAlaSerAlaSerGlnValAlaGlyThrThr 138

Qy       1103 AGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCA 1162
                |||          ||||||  ||||||||||:::   ||||||||||||:::||
Db        139 GlyVal---------CysProLeuIle-PheValPheValValGluMetGlyPheHisHi 155

Qy       1163 TGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCA 1222
              |||||||  |||||||||   ||||||   ||||||||||||      |||      ||||
Db        155 sValAla-ArgLeuValPheAsnSerSerProGlnValIleHisProProTrpProProL 175

Qy       1223 AAGTGCTGGAATTA 1236
              ||||||||   |||
Db        175 ysValLeuArgLeu 179

RESULT 13
Q8NAL9_HUMAN
ID   Q8NAL9_HUMAN PRELIMINARY;      PRT;   158 AA.
AC   Q8NAL9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ35131.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Placenta;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK092450; BAC03893.1; -; mRNA.
DR   Ensembl; ENSG00000198057; Homo sapiens.
SQ   SEQUENCE   158 AA;  17414 MW;  E151503015F2FE34 CRC64;

Alignment Scores:
Pred. No.:              3.62e-30        Length:         158
Score:                  384.50          Matches:        96
Percent Similarity:     47.2%           Conservative:   6
Best Local Similarity:  44.4%           Mismatches:     44
Query Match:            8.7%            Indels:         71
DB:                     2               Gaps:           3

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x Q8NAL9_HUMAN (1-158)

Qy        633 TCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAA 692
              |||   ||||||||||||||      |||||||||          ||||||||||||
Db          3 SerHisSerValThrGlnAlaGlyValGlnTrpCysGlyLeuGlySerLeuGlnProLeu 22

Qy        693 CCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGT 752
              ||||||      |||            |||||||||   ||||||||||||||   |||
Db         23 ProProGlyPheLysArgPhePheCysLeuSerLeuLeuSerSerTrpAspTyrArgArg 42

Qy        753 GCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTT 812
              ||||||   |||||||||
Db         43 ValProProProLeuAlaAsn---------------------------------------- 49

Qy        813 GGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGT 872
                    |||:::
Db         50 ------PheCysIle---------------------------------------------- 52

Qy        873 GCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTT 932

Db         52 ------------------------------------------------------------ 52

Qy        933 CCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTT 992
```

```
                                                  |||      |||||||||
Db         53 ----------------------------------------PhePheSerPhePhePhe 58

Qy        993 GAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGC---------------TCACT 1037
              |||   ||||||      :::|||        |||                    |||||
Db         59 GluLysGluSerLeuSerValThrGlnAlaGlyVal-GlnTrpHisAspLeuGlySerLe 78

Qy       1038 GCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAA 1097
              ||||      ||||||||||||      ||||||||||||||||||    ||||||||||
Db         78 uGlnAlaAlaProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAs 98

Qy       1098 TTACAAGCGTGCACCACCATGCTTGGCTAA-TTTTTTGTATTTTTAGCAGAGATGGGGTT 1156
              ||||:::|||   |||||||||   |||   ||||||||||||||   ||||||||||
Db         98 nTyrArgArgProProProCysProAlaAsnPhePheValPheLeuValGluMetGlyPh 118

Qy       1157 TTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGT 1216
              |   |||:::   ||||||||||||:::||||||||||||||||   |||   |||
Db        118 eCysHisIleGlyGlnAlaGlyLeuGluLeuLeuThrSerSerAspProProThrSerAl 138

Qy       1217 CTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
               |||||||||||||||||   |||:::||||||      ||||||
Db        138 aSerGlnSerAlaGlyIleThrGlyMetSerHisArgThrTrpPro 153
```

```
RESULT 14
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;
```

```
Alignment Scores:
Pred. No.:              8.67e-30        Length:         585
Score:                  382.50          Matches:        108
Percent Similarity:     56.8%           Conservative:   18
Best Local Similarity:  48.6%           Mismatches:     65
Query Match:            8.6%            Indels:         31
DB:                     1               Gaps:           2

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU5_HUMAN (1-585)

Qy       1260 GCCAGGCACAGTGACTCACGCCTATAATTCCAGCACTTTGGGAGACTGAGGCAGGCAGAT 1201
              |||          |||||||||     ||||||||||||||   |||:::||||||
Db         99 AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAsp 118

Qy       1200 CACTTGAGGTCAGGAGTTTGAGACCAGCCTGGGCAACATGGTAAAACCCCATCTCTGCTA 1141
              |||  ||||||:::      ||||||:::|||   ||||||   ||||||:::|||   |
Db        119 His--GluValLysArgSerArgProSerTrpProThrTrp***AsnProValSerThrL 138

Qy       1140 AAAATACAAAAAATTAGCCAAGCATGGTGGTGCACGCTTGTAATTCCAGCTACTCAGGAG 1081
              |||||||| |||||||||   |||||||||         |||:::||||||||:::|||
Db        138 ysAsnThr-LysIleSerTrpAlaTrpTrpArgAlaProValValProAlaThrArgGlu 157

Qy       1080 GCTTAGGCAGGAGAATCRCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCTGAACGGAGG 1021
              |||   |||||||||||||||||||||||||:::|||||||||   ||||||
Db        158 AlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGlu------ 175

Qy       1020 TTGAGGTGATGAAGTGAGACTCTGTCTCAAAAAAGAAAGGAAGAAAATAGGCAGTAATCC 961

Db        176 ----------------------------------------------------------Ile 176

Qy        960 CTGTTCTGTGGATCTTACATTCTATTGGAAAGATGGAATGTAAAGAAATAAG------TA 907
                       |||   :::      ||||||   :::|||   ||| |||
Db        177 AlaProLeuHisSerSerLeuAlaThrGluArgAspSerValSerLys-Lys******** 196

Qy        906 GCCGCACATGGTGGCTCACCCCTGTAATCCCAGCACTTTGGGATGCCAAGGTAGGAGGAT 847
               |||      |||||||||   ||||||||||||||||||||   :::|||   |||||
Db        196 *ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIl 216

Qy        846 TGCTTGAAGCCAGGAGTTTAAGACAAGCCTGGCCAACGTAGTAAGACCCAGTCTCTACAA 787
```

```
Db        216 eThrArgSerArgAspArg--AspHisProGlyGlnHisGlyGluThrProSerLeu--L 235

Qy        786 AAAAATTTTTTTAATTAGCCAGGCGTGGTGGTGCACGCTCGTAGTCCCAGCTACTTGGGA 727

Db        235 euLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyA 255

Qy        726 GGCTGAGGTCAGCAGATCCCTTGAGCTCAGGAGGTTGAGGCTGCAGTGACCTATGGTTGC 667

Db        255 rgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerA 275

Qy        666 ACCACTGTGCTCATGCCTGGGTGACAGAGCAAGACCCTGTCTCTAAACAA 617

Db        275 rgHisCysThrProAlaTrpArgGlnSerGluThrProSerGlnLysLys 291

RESULT 15
ALU3_HUMAN
ID   ALU3_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39190;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB1 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14569; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63613 MW;  85C4155726DEF235 CRC64;

Alignment Scores:
Pred. No.:              8.68e-30      Length:          587
Score:                  382.50        Matches:         141
Percent Similarity:     38.3%         Conservative:    24
Best Local Similarity:  32.7%         Mismatches:      140
Query Match:            8.6%          Indels:          126
DB:                     1             Gaps:            11

US-10-607-806-1-G7328-T9182_COPY_7000_9500 (1-2501) x ALU3_HUMAN (1-587)

Qy          2 TCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATGC 61

Db        275 SerArgHisCysThrProAlaTrpAlaThrGluArgAspSerValSerLysLys------ 292

Qy         62 TTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCCT 121

Db        293 ------------------*********PhePheLeuArgArgSerLeuAlaLeuSerPro 306

Qy        122 ------GAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGG 175

Db        307 ArgLeuGluCysSerGlyGlyIleSerAlaHisCysLysLeuArgLeuProGly------ 324

Qy        176 CTGACCTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTG 235

Db        325 ------------SerArgHisSerProAlaSerAla---------------------- 332

Qy        236 CCAGACACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCT 295

Db        333 --------------------------SerGlnValAlaGly-----------ThrThr 339

Qy        296 GGACAMMCCGTACACCCACACCTA---TTCATACTCGTGCTCTGGCTCGGCAATCACCTG 352

Db        340 GlyAlaArgHisTyrAlaArgLeuIlePheValPheLeuValGluThrGlyPheHis--- 358

Qy        353 TAGCAGTAGGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGG 412

Db        358 ------------------------------------------------------------ 358

Qy        413 GGGAAATAATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGT 472

Db        359 -----------------------------------ArgPheSerArgAspGlyLeuAsp 366

Qy        473 CTCCTTTAATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATA 532

Db        367 LeuLeuThrSer***SerAlaArgLeuGly---------------------------- 376

Qy        533 AGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGT 592
```

```
                                        |||||||||
Db        377 ----------------------------LeuProLysCysTrpAspTyrArgArgGlu 386

Qy        593 GACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGC 652
                                      |||   |||   |||   ||||||   ||||||
Db        387 ProProArgProAla*********PhePhe***AspGlyValSerLeuCysArgProGly 406

Qy        653 ATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATC 712
                ||| |||           :::||||||:::|||  |||           |||
Db        407 TrpSerAlaValAlaGlySerArgLeuThrAlaSerSerAlaSerArgValHisAlaIle 426

Qy        713 TGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAA 772
                  ||||||||||||||  |||||||||:::|||     ||||||||||||
Db        427 LeuLeuProGlnProProLys***LeuGlyLeuGlnAlaProAlaThrThrProGly*** 446

Qy        773 TTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACT 832
                    |||     |||   |||  :::  |||          :::  ::
Db        447 PheLeuTyrPhe******Arg-ArgGly---PheThrValLeuAlaGlyMetValSerIl 465

Qy        833 CCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAG 892
                :         :::||||||     ||||||||||||||||||||||||||||||
Db        465 eSer***ProArgAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSe 485

Qy        893 CCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCA 952
                ||||    |||
Db        485 rHisArgAlaArg******---------------------------------------- 491

Qy        953 CAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCAT 1012
                                              |||||||||||||||||     ::
Db        492 ------------------------------***PhePheGluThrGluSerArgSerVa 501

Qy       1013 CACCTCAACCTCCGTTCAGC--------------TCACTGCAACCTCTGCCTCCCGGGT 1057
                :         |||                |||||||||     ||||||||||
Db        501 lAlaGlnAlaGlyVal-GlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyP 521

Qy       1058 TCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCAT 1117
                ||      |||||||||||||||||   |||||||||:::|||:::|||   |||
Db        521 heThrProPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProLeuA 541

Qy       1118 GCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGG 1177
                |||||| |||||||||||||||||||||||||   |||      |||||||||
Db        541 rgProAlaAsn-PheCysIlePheSerArgAspGlyValSerProPhe***ProGlyTrp 560

Qy       1178 TCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTAT 1237
                |||::::::|||||||||  ||||||      |||:::   ||||||||||||   |||
Db        561 SerArgSerProAspLeu--ValIleArgProProArgProProLysValLeuGlyLeuG 580

Qy       1238 AGGCGTGAGTCACTGTGCCTGGC 1260
                |||      |||   ||||||
Db        580 lnAla***AlaThrAlaProGly 587


Search completed: February  8, 2006, 11:14:43
Job time : 596.638 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 16.3384 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acgggctcactctgtcgcc.....tttaataaagacggggtttc 501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_605_@runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	391	43.4	170	7	ADB64987	Adb64987 Human pro
2	381.5	42.4	382	4	AAU32610	Aau32610 Novel hum
3	381.5	42.4	382	4	AAU31818	Aau31818 Novel hum
4	381.5	42.4	382	4	AAU32707	Aau32707 Novel hum
5	378	42.0	213	4	AAM25396	Aam25396 Human pro
6	373.5	41.5	186	8	ADS11829	Ads11829 Human the
7	371.5	41.3	324	4	AAU29573	Aau29573 Novel hum
8	371.5	41.3	324	4	AAU29579	Aau29579 Novel hum
9	366.5	40.7	241	4	ABB11464	Abb11464 Human neu
10	366.5	40.7	361	5	ABG68738	Abg68738 Human pro
11	366	40.7	381	4	AAU30235	Aau30235 Novel hum
12	363	40.3	217	7	ADC87151	Adc87151 Human GPC
13	358.5	39.8	179	8	ADR09211	Adr09211 Human pro
14	355	39.4	196	7	ADE09985	Ade09985 Novel pro
15	353	39.2	296	4	AAU33304	Aau33304 Novel hum
16	345	38.3	154	4	ABG08425	Abg08425 Novel hum
17	344.5	38.3	158	7	ADB64893	Adb64893 Human pro
18	343.5	38.2	302	5	ADK36936	Adk36936 Novel hum
19	342	38.0	203	7	ADM04063	Adm04063 Human pro
20	341.5	37.9	257	4	AAO10294	Aao10294 Human pol
21	338.5	37.6	201	7	ADC86633	Adc86633 Human GPC
22	338.5	37.6	411	4	ABG08428	Abg08428 Novel hum
23	329.5	36.6	154	6	ADA54692	Ada54692 Human pro
24	328.5	36.5	293	4	AAU31676	Aau31676 Novel hum
25	328	36.4	158	4	AAU30493	Aau30493 Novel hum
26	321	35.7	218	7	ADC86939	Adc86939 Human GPC
27	316.5	35.2	185	4	AAO06014	Aao06014 Human pol
28	314.5	34.9	375	3	AAB01399	Aab01399 Neuron-as
29	314.5	34.9	375	5	ABB81538	Abb81538 Neural th
30	314.5	34.9	375	5	AAE29142	Aae29142 AD7c-NTP
31	314.5	34.9	375	6	ABR63268	Abr63268 AD7c-NTP
32	314.5	34.9	375	6	ADA84017	Ada84017 Human POM
33	314.5	34.9	375	6	ABU03024	Abu03024 Human neu
34	314.5	34.9	375	6	ABB99774	Abb99774 Amino aci
35	314.5	34.9	375	6	ABJ19445	Abj19445 AD7C-neur
36	314.5	34.9	375	7	ADB37642	Adb37642 Human neu
37	314.5	34.9	375	8	ADR14409	Adr14409 Human NF-
38	314.5	34.9	375	9	AEA30020	Aea30020 Human neu
39	312	34.7	603	7	ADJ68334	Adj68334 Human hea
40	311	34.6	239	8	ADQ65417	Adq65417 Novel hum
41	305	33.9	215	4	AAU31513	Aau31513 Novel hum
42	300	33.3	167	7	ADC86521	Adc86521 Human GPC
43	299.5	33.3	189	7	ADB65042	Adb65042 Human pro
44	298.5	33.2	290	4	ABG21913	Abg21913 Novel hum
45	298	33.1	146	4	ABG08195	Abg08195 Novel hum

ALIGNMENTS

RESULT 1
ADB64987
ID ADB64987 standard; protein; 170 AA.
XX
AC ADB64987;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone PROST20121570.
XX
KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX
OS Homo sapiens.
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX

```
DR   WPI; 2003-450961/43.
DR   N-PSDB; ADB63017.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a protein of the invention. Note: Some of the
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 170 AA;

Alignment Scores:
Pred. No.:                3.11e-37      Length:          170
Score:                    391.00        Matches:         96
Percent Similarity:       60.8%         Conservative:    14
Best Local Similarity:    53.0%         Mismatches:      34
Query Match:              43.4%         Indels:          37
DB:                       7             Gaps:            5

US-10-607-806-1_COPY_5000_5500 (1-501) x ADB64987 (1-170)

Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::|||   |||   ||||||||||||||||||||||   |||   ||||||||||||||
Db          9 SerHisCysValThrGlnAlaGlyValArgTrpArgGlyLeuSerSerLeuGlnProPro 28

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125
              ||||||||||||:::   ||||||||||||   ||| ||||||:::|||
Db         29 SerProGlyPheGlnArgPheSerCysLeuGlyLeu-GlnValSerGly----------- 44

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
                 :::||||||   ||||||||||||||||||||||||        ||||||||||||
Db         45 ----IleAlaGlyAlaHisHisHisAlaGlnLeuIlePhe------ValPheLeuValGl 61

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCT 245
              |   ||||||||||||||||||||||||||||                  |||   :::||
Db         61 uMetGlyPheHisHisValGlyGlnAlaGly-----------------LeuGluLeuLe 75

Qy        246 GCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCA----------- 294
              |      ::::::|||                    |||   |||:::
Db         75 uThrSerAlaAsnPro--------------------ProSerLeuAlaSerArgGlyAl 88

Qy        295 ------------------------------CATTTATTATTATTCTTATTATTATTGAG 323
                                           |||           :::   |||:::||
Db         88 aGlyIleThrGlyMetSerHisLeuThrGlnHisPhePhePheLeuIlePheLeuMetAr 108

Qy        324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
              ||||||||||:::||||||||||||||||||   ||||||   ||||||||||||||||
Db        108 gGlnSerLeuSerLeuSerProArgLeuGluCysSerGlyThrIleSerAlaHisCysAs 128

Qy        384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
              |||   |||      |||   :::||||||   ||||||      ||||||||||||
Db        128 nLeuCysLeuProCysSerSerAsnSerProThrSerAlaTyrTrpValAlaGlyThrTh 148

Qy        444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTTT 500
              |||      |||||||||   |||     ||||||||||||::::::   ||||||
Db        148 rGlyValHisHisHisAlaGlnLeu--SerPheValPheLeuValGluMetGlyPhe 166

RESULT 2
AAU32610
ID   AAU32610 standard; protein; 382 AA.
XX
AC   AAU32610;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #3101.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 637; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 382 AA;

Alignment Scores:
Pred. No.:                5.39e-36      Length:          382
Score:                    381.50        Matches:         90
Percent Similarity:       56.6%         Conservative:    13
Best Local Similarity:    49.5%         Mismatches:      43
Query Match:              42.4%         Indels:          36
DB:                       4             Gaps:            6
```

```
US-10-607-806-1_COPY_5000_5500 (1-501) x AAU32610 (1-382)

Qy          8 TCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATC 67
              |||||||||   ||||||||||||   ||||||   |||:::||||||   |||||| ::
Db        103 SerLeuCysTyrProGlyTrpSerArgValAla***SerArgIleThrAlaThrSer-Al 122

Qy         68 TCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTC 127
              :||||||   |||   |||:::|||   ||||||   ||||||   |||
Db        122 aProGlyLeuLys***PheAlaCysPheSerLeuProSerSerArgAspTyr-------- 139

Qy        128 AAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGA 187
                  |||   ||||||   |||   ||||||   |||:::   ||||||||||
Db        140 -------ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyArgAs 155

Qy        188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGC 247
              |      ||||||||||||||||||||      :::                   |||
Db        155 pGluValSerProCysTrpProGlyTrpPhe***Thr------------------ProAs 169

Qy        248 TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTA-------------------- 287
               |||   :::|||                |||   |||
Db        169 pLeuArgTyrPro--------------ProAlaSerAlaSerGlnSerAlaGluIleIl 184

Qy        288 ------------------------------CTTATCACATTTATTATTATTCTTATTAT 316
                                           :::      |||:::   :::   |||
Db        184 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIleTy 204

Qy        317 TATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTC 376
               :::                |||||:::|||   ||||||      ||||||||||
Db        204 rLeuArgTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgArgAspLeuGlySe 224

Qy        377 ACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTG 436
              ||||||||||||||||   |||   |||||||||   ||||||||||||||||||||
Db        224 rLeuGlnAlaProProProArgPheLysProPheSerCysLeuSerLeuProSerSerTr 244

Qy        437 GGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGG 496
              ||||   ||||||   ||||||   ||||||||||||||||||||||:::::::|||||
Db        244 pAspTyrArgArgProProProHisProAlaAsnPhePheValPheLeuValGluThrGl 264

Qy        497 GTTT 500
              ||||
Db        264 yPhe 265

RESULT 3
AAU31818
ID   AAU31818 standard; protein; 382 AA.
XX
AC   AAU31818;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #2309.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 513; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 382 AA;

Alignment Scores:
Pred. No.:              5.39e-36        Length:         382
Score:                  381.50          Matches:        90
Percent Similarity:     56.6%           Conservative:   13
Best Local Similarity:  49.5%           Mismatches:     43
Query Match:            42.4%           Indels:         36
DB:                     4               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU31818 (1-382)

Qy          8 TCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATC 67
              |||||||||   ||||||||||||   ||||||   |||:::||||||   |||||| ::
Db        103 SerLeuCysTyrProGlyTrpSerArgValAla***SerArgIleThrAlaThrSer-Al 122

Qy         68 TCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTC 127
              :||||||   |||   |||:::|||   ||||||   ||||||   |||
Db        122 aProGlyLeuLys***PheAlaCysPheSerLeuProSerSerArgAspTyr-------- 139

Qy        128 AAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGA 187
                  |||   ||||||   |||   ||||||   |||:::   ||||||||||
Db        140 -------ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyArgAs 155

Qy        188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGC 247
              |      ||||||||||||||||||||      :::                   |||
Db        155 pGluValSerProCysTrpProGlyTrpPhe***Thr------------------ProAs 169

Qy        248 TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTA-------------------- 287
               |||   :::|||                |||   |||
Db        169 pLeuArgTyrPro--------------ProAlaSerAlaSerGlnSerAlaGluIleIl 184

Qy        288 ------------------------------CTTATCACATTTATTATTATTCTTATTAT 316
                                           :::      |||:::   :::   |||
Db        184 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIleTy 204

Qy        317 TATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTC 376
               :::                |||||:::|||   ||||||      ||||||||||
Db        204 rLeuArgTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgArgAspLeuGlySe 224

Qy        377 ACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTG 436
              ||||||||||||||||   |||   |||||||||   ||||||||||||||||||||
Db        224 rLeuGlnAlaProProProArgPheLysProPheSerCysLeuSerLeuProSerSerTr 244

Qy        437 GGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGG 496
              ||||   ||||||   ||||||   ||||||||||||||||||||||:::::::|||||
Db        244 pAspTyrArgArgProProProHisProAlaAsnPhePheValPheLeuValGluThrGl 264

Qy        497 GTTT 500
              ||||
```

```
Db        264 yPhe 265

RESULT 4
AAU32707
ID   AAU32707 standard; protein; 382 AA.
XX
AC   AAU32707;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #3198.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 652; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 382 AA;

Alignment Scores:
Pred. No.:              5.39e-36      Length:          382
Score:                  381.50        Matches:         90
Percent Similarity:     56.6%         Conservative:    13
Best Local Similarity:  49.5%         Mismatches:      43
Query Match:            42.4%         Indels:          36
DB:                     4             Gaps:            6

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU32707 (1-382)

Qy          8 TCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATC 67
              |||||||||   ||||||||||||   ||||||   |||:::||||||   |||||| ::
Db        103 SerLeuCysTyrProGlyTrpSerArgValAla***SerArgIleThrAlaThrSer-Al 122

Qy         68 TCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTC 127
                :||||||   |||   |||:::|||   ||||||   ||||||   |||
Db        122 aProGlyLeuLys***PheAlaCysPheSerLeuProSerSerArgAspTyr-------- 139

Qy        128 AAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGA 187
                 |||   ||||||   |||   ||||||   |||:::   ||||||||||
Db        140 -------ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyArgAs 155

Qy        188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGC 247
              |   ||||||||||||||||||||||      :::                  |||
Db        155 pGluValSerProCysTrpProGlyTrpPhe***Thr-----------------ProAs 169

Qy        248 TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTA-------------------- 287
               |||   :::|||               |||   |||
Db        169 pLeuArgTyrPro--------------ProAlaSerAlaSerGlnSerAlaGluIleIl 184

Qy        288 ------------------------------CTTATCACATTTATTATTATTCTTATTAT 316
                                           :::      |||:::   :::   |||
Db        184 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIleTy 204

Qy        317 TATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTC 376
               :::              ||||||:::|||   ||||||      ||||||||||
Db        204 rLeuArgTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgArgAspLeuGlySe 224

Qy        377 ACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTG 436
              ||||||||||||||||      |||   |||||||||   ||||||||||||||||||
Db        224 rLeuGlnAlaProProProArgPheLysProPheSerCysLeuSerLeuProSerSerTr 244

Qy        437 GGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGG 496
              ||||   ||||||   ||||||   ||||||||||||||||||||||:::::|||||
Db        244 pAspTyrArgArgProProProHisProAlaAsnPhePheValPheLeuValGluThrGl 264

Qy        497 GTTT 500
              ||||
Db        264 yPhe 265

RESULT 5
AAM25396
ID   AAM25396 standard; protein; 213 AA.
XX
AC   AAM25396;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:911.
XX
KW   Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW   antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW   antibacterial; endocrine; cardiant; central nervous system; virucide;
KW   anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW   antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW   dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW   neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW   immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW   antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW   cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW   genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW   thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW   allergic rhinitis; diabetes; multiple sclerosis; depression;
KW   Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW   neurological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200153455-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US035017.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
XX
```

```
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457603/49.
DR   N-PSDB; AAH99337.
XX
PT   Isolated human polynucleotides encoding polypeptides, useful for the
PT   treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
PS   Claim 20; Page 198; 1217pp; English.
XX
CC   AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC   AAM25963. The proteins can have activities based on the tissues and cells
CC   they are expressed in, such as: antiinflammatory; antirheumatic;
CC   antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC   central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC   cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
CC   antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC   antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC   antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC   encoding them can be used in gene therapy, antisense therapy and vaccine
CC   production, The proteins and polynucleotides are useful for screening for
CC   agonists or antagonists of a protein and for the treatment and diagnosis
CC   of disorders associated with the activity of a protein e.g. inflammation,
CC   rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC   neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC   infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC   anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC   osteoporosis, severe combined immunodeficiency, eczema, allergic
CC   rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC   Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC   neurological disorders
XX
SQ   Sequence 213 AA;

Alignment Scores:
Pred. No.:                1.18e-35        Length:          213
Score:                    378.00          Matches:         94
Percent Similarity:       63.6%           Conservative:    11
Best Local Similarity:    57.0%           Mismatches:      41
Query Match:              42.0%           Indels:          20
DB:                       4               Gaps:            4

US-10-607-806-1_COPY_5000_5500 (1-501) x AAM25396 (1-213)

Qy         12 TCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCT 71
              ||||||||||||||||   |||:::|||   ||||||||||||||||||   |||   |||
Db         37 SerValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaProProPro 56

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
              ||||||      ||||||||||||||||||      |||||||||
Db         57 GlyPheThrProPheSerCysLeuSerLeuProGlySerTrpAspTyrArgCys------ 74

Qy        132 AGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAG-AGACAG 190
                 ||||||   ||||||||||||||||      :::|||      |||:::
Db         75 ---------ProProProArgProAlaAsnPheLeu------TyrPhe******ArgArg 89

Qy        191 GTTTTCACCATGTTGGCCAGGCTGGGTGACCTC------CCTTTTAGATTCTCCTCATCC 244
                ||||||:::|||||||||:::      :::      |||     |||   |||
Db         90 GlyPheThrValLeuAlaArgMetValSerIleSer***Pro-ArgAspProProAlaSe 109

Qy        245 TGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTAT 304
              |||                                |||      ||||||   :::
Db        109 rAla--------------------------SerGlnSerAlaGlyIleThrValLeuSe 120

Qy        305 TATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTG 364
               :::             |||   |||||||||   ||||||:::|||||||||||||||
Db        120 rLeuPhePhePhePheGluMetGluSerCysSerValAlaGlnAlaGlyValGlnTrpAr 140

Qy        365 CGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCT 424
              ||||||||||||||||   |||   ||||||||||||||||||   ||||||||
Db        140 gTyrLeuGlySerLeuGlnAlaLeuProProGlyPheThrProPheSerCysLeuSerLe 160

Qy        425 CCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTT 484
              ||||||||||||||||   ||||||   ||||||||||||||||||||||||||||||
Db        160 uProSerSerTrpAspTyrArgArgProProProArgProAlaAsnPhePheValPheLe 180

Qy        485 AATAAAGACGGGG 497
              |:::::|||||
Db        180 uValGluThrGly 184

RESULT 6
ADS11829
ID   ADS11829 standard; protein; 186 AA.
XX
AC   ADS11829;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human therapeutic contig protein - SEQ ID 2066.
XX
KW   antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;
KW   inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
KW   aplastic anaemia; cancer; wound healing; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .186
FT                   /label= Unknown, OTHER
FT                   /note= "OTHER = In-frame STOP codon"
XX
PN   WO2004080148-A2.
XX
PD   23-SEP-2004.
XX
PF   30-SEP-2003; 2003WO-US030720.
XX
PR   02-OCT-2002; 2002US-0416186P.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Tang YT,  Asundi V,  Ren F,  Zhang J,  Wehrman T,  Wang Z,  Ma Y;
PI   Wang D,  Chen R,  Zhao QA,  Wang J,  Ghosh M,  Xue AJ,  Weng G,  Zhou P;
XX
DR   WPI; 2004-668857/65.
DR   N-PSDB; ADS11231.
XX
PT   New polynucleotide, useful in preparing a composition for diagnosing or
PT   treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
PT   aplastic anemia or cancer for promoting wound healing.
XX
PS   Example 2; SEQ ID NO 2066; 718pp; English.
XX
CC   The invention relates to a novel isolated polynucleotide and the encoded
CC   polypeptide. The molecules of the invention demonstrate antiinflammatory,
CC   neuroprotective, antianaemic, cytostatic and vulnerary activities and may
CC   be useful in preparing a composition for diagnosing or treating
CC   inflammatory, haematopoietic, immune, neurodegenerative or stem cell
CC   disorders, such as aplastic anaemia or cancer, as well as for promoting
CC   wound healing. The molecules may also be utilised during gene therapy
CC   procedures. The current sequence is that of a human therapeutic contig
CC   protein of the invention.
XX
SQ   Sequence 186 AA;

Alignment Scores:
Pred. No.:                3.91e-35        Length:          186
Score:                    373.50          Matches:         93
Percent Similarity:       60.5%           Conservative:    11
Best Local Similarity:    54.1%           Mismatches:      51
Query Match:              41.5%           Indels:          17
DB:                       8               Gaps:            4

US-10-607-806-1_COPY_5000_5500 (1-501) x ADS11829 (1-186)
```

```
Qy         12 TCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCT 71
Db         16 SerValAlaGlnAlaGluValGlnTrpProHisLeuAsnSerLeuGlnAlaProProPro 35

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
Db         36 GlyPheAlaProPheSerCysLeuArgLeuProSerSerTrpAspTyr------------ 51

Qy        132 AGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGG 191
Db         52 ---ArgHisLeuProProCysProAlaAsnPheLeu--TyrPheTrpTrp---ArgArgG 69

Qy        192 TTTTCACCATGTTGGCCAGGCTG----------------GGTGACCTCCCTTTTAGAT 233
Db         69 lyPheThrMetLeuAlaArgMetValLeuIleSer***ProArgAspProProAlaSerA 89

Qy        234 TCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATC 293
Db         89 laSerGlnGlyAlaGlyIleAlaGlyMetSerHisCysAlaArgPro***MetAsnTyr- 108

Qy        294 ACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGA 353
Db        109 --PheTyrLeuPheIleTyrPhePhe-GluMetGluSerArgSerValAlaGlnAlaGlu 127

Qy        354 GTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCC 413
Db        128 ValGlnTrpProHisLeuAsnSerLeuGlnAlaProProProGlyPheAlaProPheSer 147

Qy        414 CGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTT 473
Db        148 CysLeuArgLeuProSerSerTrpAspTyrArgHisLeuProProCysProAlaAsn-Ph 167

Qy        474 TTTGTATTTTTAATAAAGACGGGGTTTC 501
Db        167 eLeuTyrPheTrpTrpArgArgGlyPhe 176
```

RESULT 7
AAU29573
ID AAU29573 standard; protein; 324 AA.
XX
AC AAU29573;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #64.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 167; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 324 AA;

Alignment Scores:

Pred. No.:	7.99e-35	Length:	324
Score:	371.50	Matches:	95
Percent Similarity:	60.4%	Conservative:	15
Best Local Similarity:	52.2%	Mismatches:	44
Query Match:	41.3%	Indels:	28
DB:	4	Gaps:	5

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU29573 (1-324)

```
Qy         15 GTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAA---CCTCCATCTCCT 71
Db        140 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProPro 159

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
Db        160 ArgPheLysGlnPheSerCysLeuSerLeuProSerThrTrp------------------ 173

Qy        132 AGCTGG---------CACACACCACCATGCCCAGCTAATTTTTTGTGTG----------- 171
Db        174 SerTrpValTyrArgHisAlaProProCysProAlaAsnPheVal-PhePhePhePhPh 193

Qy        172 ----TTTTTTTTGGTAGAGACAGGTTTTCACCATGTTGGCCAGGCTG------------- 214
Db        193 ePhePhePheLeuValGluLysGly-PheSerMetLeuLeuArgLeuValLeuAsnSer* 213

Qy        215 -----------GGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCCCTTT 263
Db        213 **ProHisAspProAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValS 233

Qy        264 CTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTATTGAG 323
Db        233 erHisHisThrArgProMetSerPheLysAsnIleTyrPhePhePhePhePhe-Glu 252

Qy        324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
Db        253 ThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuSerSerArgGln 272

Qy        384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
Db        273 ProProProProGlyPheLysArgPheSerCysLeuSerLeuSerSerSerTrpAspTyr 292

Qy        444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
Db        293 ArgArgValProProMetProGly--***PheCysIlePheArgArgAspGlyVal 310
```

RESULT 8
AAU29579
ID AAU29579 standard; protein; 324 AA.
XX
AC AAU29579;
XX
DT 18-DEC-2001 (first entry)

```
XX
DE   Novel human secreted protein #70.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 168; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 324 AA;

Alignment Scores:
Pred. No.:              7.99e-35      Length:          324
Score:                  371.50        Matches:         95
Percent Similarity:     60.4%         Conservative:    15
Best Local Similarity:  52.2%         Mismatches:      44
Query Match:            41.3%         Indels:          28
DB:                     4             Gaps:            5

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU29579 (1-324)

Qy         15 GTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAA---CCTCCATCTCCT 71
              ||||||||||||||||||:::||||||||||||   |||||||||   ||||||   |||
Db        140 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProPro 159

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
                ||||||   ||||||||||||||||||   |||:::|||
Db        160 ArgPheLysGlnPheSerCysLeuSerLeuProSerThrTrp---------------------- 173

Qy        132 AGCTGG---------CACACACCACCATGCCCAGCTAATTTTTTGTGTG---------- 171
              ||||||         |||   ||||||||||||||||||||:::
Db        174 SerTrpValTyrArgHisAlaProProCysProAlaAsnPheVal-PhePhePhePhePh 193

Qy        172 ----TTTTTTTTGGTAGAGACAGGTTTTCACCATGTTGGCCAGGCTG------------- 214
                  ||||||||||||||   ||| |||:::||||||   ||||||

Db        193 ePhePhePheLeuValGluLysGly-PheSerMetLeuLeuArgLeuValLeuAsnSer* 213

Qy        215 -----------GGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCCCTTT 263
                          |||   |||         |||   |||     |||      :::|
Db        213 **ProHisAspProAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValS 233

Qy        264 CTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTATTGAG 323
              ||:::                    :::||||||::::::   ::::::::: |||
Db        233 erHisHisThrArgProMetSerPheLysAsnIleTyrPhePhePhePhePhe-Glu 252

Qy        324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
              |||||||||      ||||||:::|||||||||||||||   ||||||   |||   |||
Db        253 ThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuSerSerArgGln 272

Qy        384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
                ||||||   ||||||      ||||||   |||||||||   ||||||||||||
Db        273 ProProProProGlyPheLysArgPheSerCysLeuSerLeuSerSerSerTrpAspTyr 292

Qy        444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
              ||||||:::||||||   |||        ||||||||||||   :::|||||||||
Db        293 ArgArgValProProMetProGly--***PheCysIlePheArgArgAspGlyVal 310

RESULT 9
ABB11464
ID   ABB11464 standard; peptide; 241 AA.
XX
AC   ABB11464;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human neuronal thread protein homologue, SEQ ID NO:1834.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US003800.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   N-PSDB; ABA08708.
XX
PT   Human proteins and DNA encoding sequences useful for preventing, treating
PT   or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT   and cancer.
XX
PS   Claim 20; Page 194; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
```

```
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention
XX
SQ   Sequence 241 AA;

Alignment Scores:
Pred. No.:              2.89e-34        Length:         241
Score:                  366.50          Matches:        96
Percent Similarity:     52.1%           Conservative:   17
Best Local Similarity:  44.2%           Mismatches:     44
Query Match:            40.7%           Indels:         60
DB:                     4               Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x ABB11464 (1-241)

Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::   |||||||||||||||||||||:::|||   :::|||   |||||||||
Db          8 SerArgSerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnAlaLeu 27

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125
                 |||||||| ::::::|||   ||||||||||||:::||||||||||||
Db         28 ProProGlyPhe-SerHisSerProAlaSerAlaSerArgValAlaGlyThr-------- 44

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
                   ||||||   |||   |||:::|||||||||     :::   |||   |||
Db         45 -------ThrGlyThrArgHis***AlaArgLeuIlePhe--TyrIle---PheSerArg 60

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCC---------------------- 224
                 |||       |||||||||   |||||||||       |||
Db         61 AspGlyValSerProCys***ProGlyTrpSer***SerProAspLeuValIleArgPro 80

Qy        225 --------------------------CTTTTAGATTCTCCTCATCCTGCTCTATTCTTC 257
                                       :::   |||   ||||||   ||||||
Db         81 ProArgLeuProLysCysTrpAspTyrArgArgGluProProArgProAla***PhePhe 100

Qy        258 CCCTTTCTAATGCAGTATCCAGTTTCCTTACTT-------------------------- 290
                 ||||||::::::         ::::::|||
Db        101 ValPheLeuValGluGlnGlyPheThrMetLeuAlaArgMetValSerIleSer***Pro 120

Qy        290 ------------------------------------------------------------ 290

Db        121 GlnCysAspLeuProAlaSerValSerGlnAsnAlaGlyIleThrGlyValSerHisCys 140

Qy        291 ------------ATCACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTT 338
                       :::   |||                     |||   |||||||||
Db        141 AlaTrpProCysLeuHisPheCysPhePheGlyPhePhePheGluMetGluSerCysSer 160

Qy        339 GTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGG 398
                 ||||||:::|||   |||||||||   ||||||   ||||||||||||||||||   |||
Db        161 ValAlaGlnAlaGluValGlnTrpHisAspLeuArgSerLeuGlnAlaProProProGly 180

Qy        399 TTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCA 458
                 ||||||||||||||   ||||||||||||   |||||||||   ||||||   ||||||
Db        181 PheThrProPheSerCysLeuSerLeuProGlySerTrpAspTyrArgArgProProPro 200

Qy        459 CGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
                 ||||||||||||  ||||||||||||::::::|||||||||
Db        201 ArgProAlaAsn--PheCysIlePheSerArgAspGlyVal 213

RESULT 10
ABG68738
ID   ABG68738 standard; protein; 361 AA.
XX
AC   ABG68738;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Human prostate specific protein DEX0293_107.
XX
KW   Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
KW   prostate specific protein; cytostatic; non-cancerous prostate disease;
KW   gene therapy; cancer; immunostimulant; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200255735-A2.
XX
PD   18-JUL-2002.
XX
PF   27-NOV-2001; 2001WO-US044363.
XX
PR   27-NOV-2000; 2000US-0253176P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Recipon H,  Cafferkey R,  Ali S,  Sun Y;
PI   Liu C,  Chen S;
XX
DR   WPI; 2002-557831/59.
DR   N-PSDB; ABK97631.
XX
PT   New prostate specific genes, useful for treating or diagnosing cancer, or
PT   useful as vaccines for treating cancer, particularly prostate cancer, in
PT   a patient.
XX
PS   Claim 11; Page 207-208; 212pp; English.
XX
CC   The invention relates to a new isolated prostate-specific nucleic acid
CC   (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
CC   ABK97642 which encode prostate specific proteins appearing as ABG68701-
CC   ABG68746, or a sequence hybridising to a PSNA or which has 60% sequence
CC   homology with a PSNA. Also included are a method of determining the
CC   presence of a PSNA in a sample, a vector comprising the PSNA, a host cell
CC   comprising the vector, producing the polypeptide encoded by the PSNA, a
CC   method of determining the presence of a PSP in a sample, diagnosing and
CC   monitoring the presence and metastases of prostate cancer in a patient, a
CC   kit for detecting a risk of cancer or presence of cancer in a patient
CC   (the kit comprising a means for determining the presence of the PSNA or
CC   PSP in a sample of a patient) and a vaccine comprising the polypeptide or
CC   the nucleic acid encoding the polypeptide. The PSNA, PSP and anti-PSP
CC   antibody are useful for diagnosing and treating cancer in a patient (e.g.
CC   by gene therapy). The nucleic acid molecule and polypeptide are also
CC   useful as vaccines for treating cancer, particularly prostate cancer and
CC   non-cancerous prostate diseases. The present sequence is a PSP of the
```

```
CC   invention
XX
SQ   Sequence 361 AA;

Alignment Scores:
Pred. No.:              3.26e-34        Length:         361
Score:                  366.50          Matches:        96
Percent Similarity:     60.4%           Conservative:   17
Best Local Similarity:  51.3%           Mismatches:     44
Query Match:            40.7%           Indels:         30
DB:                     5               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x ABG68738 (1-361)

Qy          9 CACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCT 68
              ||||||||||||||||||||||||:::|||   :::|||   ||||||||||||
Db        171 HisSerValAlaGlnAlaGlyValGlnTrpCysAsnLeuGlySerLeuGlnProLeuPro 190

Qy         69 CCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCA 128
              |||   |||:::   ||||||||||||||||||   ||||||||||||
Db        191 ProArgPheArgArgPheSerCysLeuSerLeuLeuSerSerTrpAsp------------ 206

Qy        129 AGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTT---TTTTTGGTAGA 185
                  |||   |||||||||   ||||||||||||    |||   |||||||||::
Db        207 ---HisArgHisAlaProProCysLeuAlaAsnPheLeu-PhePheLysPheLeuValAs 225

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTG--------------------GGTGACCTCC 224
              : |||   ||||||||||||||||||||                    |||||||||
Db        225 p-GlnSerPheThrMetLeuAlaArgLeuValLeuAsnSerAlaProSerGlyAspLeuP 245

Qy        225 CTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTA------TCCA 278
              ||         |||   |||      |||      :::   :::               |
Db        245 roAlaProAlaSerGlnSerAlaGlyIleThrGlyValArgHisCysThrTrpProLysS 265

Qy        279 GTTTC-----------------------CTTACTTATCAC------ATTTATTATTATT 308
              |||||                      |||   :::|||         :::::::::
Db        265 erPheLysPheAlaAspSerHisIleGlyLeuAlaPheHisPheAlaPhePhePhePheP 285

Qy        309 CTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGAT 368
                :::::::::          |||      :::|||:::|||||||||||||||   |||
Db        285 hePhePhePhe-AlaValAlaSerHisProIleAlaGlnAlaGlyValGlnTrpArgAsp 304

Qy        369 CTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCC 428
              ||||||||||||||   ||||||   ||||||      |||      ||||||||||||
Db        305 LeuGlySerLeuGlnProProProProGlyPheLysGlnPheLeuCysLeuSerLeuPro 324

Qy        429 AGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATA 488
              |||||||||   ||||||   |||||||||   ||||||  ||||||||||||::::
Db        325 GlySerTrpAspTyrArgArgAlaProProArgGlnAlaAsn--PheCysIlePheSerA 344

Qy        489 AAGACGGGGTT 499
              ::|||||||||
Db        344 rgAspGlyVal 347

RESULT 11
AAU30235
ID   AAU30235 standard; protein; 381 AA.
XX
AC   AAU30235;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #726.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US008656.
XX
PR   18-APR-2000; 2000US-00552929.
PR   26-JAN-2001; 2001US-00770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy.
XX
PS   Claim 20; Page 264-265; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising the
CC   nucleic acids encoding the polypeptides and cells genetically engineered
CC   to express them are also useful for producing the proteins. The proteins
CC   are useful in genetic vaccination, testing and therapy, and can be used
CC   as nutritional supplements. They may be used to increase stem cell
CC   proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC   and/or nerve tissue growth or regeneration; immune suppression and/or
CC   stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC   AAU29510-AAU33304 represent the amino acid sequences of novel human
CC   secreted proteins of the invention
XX
SQ   Sequence 381 AA;

Alignment Scores:
Pred. No.:              3.8e-34         Length:         381
Score:                  366.00          Matches:        89
Percent Similarity:     56.0%           Conservative:   13
Best Local Similarity:  48.9%           Mismatches:     43
Query Match:            40.7%           Indels:         37
DB:                     4               Gaps:           7

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU30235 (1-381)

Qy          8 TCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATC 67
              |||||||||   ||||||||||||   ||||||   |||:::|||   |||   ::: ||
Db        103 SerLeuCysTyrProGlyTrpSerArgValAla***SerArgIle---ThrAlaThr-Se 121

Qy         68 TCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTC 127
              |||||||   |||   |||:::|||   ||||||   ||||||   |||
Db        121 rProGlyLeuLys***PheAlaCysPheSerLeuProSerSerArgAspTyr-------- 138

Qy        128 AAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGA 187
                  |||   ||||||   |||   ||||||   |||:::   ||||||||||
Db        139 -------ArgHisValProProHisProGlyAsnPhe---CysIle---PheGlyArgAs 154

Qy        188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGC 247
              |      ||||||||||||||||||||      :::                   |||
Db        154 pGluValSerProCysTrpProGlyTrpPhe***Thr-----------------ProAs 168

Qy        248 TCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTA-------------------- 287
                |||   :::|||                  |||   |||
Db        168 pLeuArgTyrPro---------------ProAlaSerAlaSerGlnSerAlaGluIleIl 183

Qy        288 ------------------------------CTTATCACATTTATTATTATTCTTATTAT 316
                                           :::      |||:::   :::   |||
Db        183 eGlyValSerHisHisThrTrpProGlnGluValPheLeuPheLeuAsnLeuPheIleTy 203

Qy        317 TATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTC 376
                :::                  |||||||:::|||   ||||||      ||||||||||
```

```
Db        203 rLeuArgTrpSerLeuAspSerValAlaGlnAlaArgValGlnArgArgAspLeuGlySe 223

Qy        377 ACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTG 436
              ||||||||||||||||      |||   |||||||||   ||||||||||||||||||||
Db        223 rLeuGlnAlaProProProArgPheLysProPheSerCysLeuSerLeuProSerSerTr 243

Qy        437 GGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGG 496
              ||||   ||||||   ||||||   ||||||||||||||||||||||::::::|||||
Db        243 pAspTyrArgArgProProProHisProAlaAsnPhePheValPheLeuValGluThrGl 263

Qy        497 GTTT 500
              ||||
Db        263 yPhe 264

RESULT 12
ADC87151
ID   ADC87151 standard; protein; 217 AA.
XX
AC   ADC87151;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR protein SEQ ID NO:1604.
XX
KW   human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   N-PSDB; ADC87150.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 2; SEQ ID NO 1604; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The protein
CC   sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ   Sequence 217 AA;

Alignment Scores:
Pred. No.:              7.32e-34      Length:         217
Score:                  363.00        Matches:        96
Percent Similarity:     55.2%         Conservative:   5
Best Local Similarity:  52.5%         Mismatches:     38
Query Match:            40.3%         Indels:         44
DB:                     7             Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x ADC87151 (1-217)

Qy          7 CTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCAT 66
              |||||||||   ||||||||||||||   |||   ||||||   |||   |||||||||

Db         55 LeuThrLeuLeuProArgLeuGluCysSerGlyThrIleSerAlaHisCysAsnLeuHis 74

Qy         67 CTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCT 126
              |||   ||||||   :::|||      ||||||||||||||||||||
Db         75 LeuProGlySerGlyAspSerHisThrSerAlaSerGlnValAlaGly------------ 90

Qy        127 CAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAG 186
                 :::   |||   |||   ||||||   |||||||||         ||||||:::|||
Db         91 ---IleThrGlyAlaHisArgHisAlaTrpLeuIlePhe------ValPheLeuLeuGlu 107

Qy        187 ACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCT---------- 236
              |||||||||   ||||||||||||||||         ||||||   |||
Db        108 ThrGlyPheCysHisValGlyGlnAlaGly--LeuGluLeuLeuThrSerGlyAspProP 127

Qy        237 ------------------------------------CCTCATCCTGCTCTATTCTTCC 258
                                                   ||||||      ||||||
Db        127 roAlaLeuAlaSerGlnSerAlaGlyIleThrGlyValSerHisProLysProPhePheP 147

Qy        259 CCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTA 318
                 |||
Db        147 hePhe-------------------------------------------------PheP 150

Qy        319 TTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCAC 378
              ||||||||||||      |||   ||||||||||||||||||||||||||||||||
Db        150 heGluThrGluSerHisThrValThrTrpAlaGlyValGlnTrpCysAspLeuGlySerL 170

Qy        379 TGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGG 438
              |||||   ||||||   ||||||      ||||||||||||||||   |||||||||
Db        170 euGlnProProProPro-GlySerSerAspSerProAlaSerAlaSerArgValAlaGly 189

Qy        439 ACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGT 498
              |||||||   |||||||||   ||||||||||||::: |||   :::|||||||||
Db        190 IleThrGlyAlaHisHisHisAlaGlnLeuIlePheLeuPhe-PheArgArgAspGlyVa 209

Qy        499 T 499
              |
Db        209 l 209

RESULT 13
ADR09211
ID   ADR09211 standard; protein; 179 AA.
XX
AC   ADR09211;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Human protein useful for treating neurological disease Seq 2717.
XX
KW   human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   N-PSDB; ADR07255.
```

```
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 2717; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polypeptide is a protein
CC   encoded by a full length human cDNA sequence of the invention. NOTE: This
CC   sequence is not given in the sequence listing of the specification but
CC   can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC   office.
XX
SQ   Sequence 179 AA;

Alignment Scores:
Pred. No.:              2.38e-33      Length:          179
Score:                  358.50        Matches:         92
Percent Similarity:     61.3%         Conservative:    11
Best Local Similarity:  54.8%         Mismatches:      43
Query Match:            39.8%         Indels:          22
DB:                     8             Gaps:            5

US-10-607-806-1_COPY_5000_5500 (1-501) x ADR09211 (1-179)

Qy         12 TCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCT 71
              ||||||:::|||||||||||:::|||   ||||||   ||||||||||||||
Db          5 SerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProProProSer 24

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
              |||||||||  ||||||||||||||||||   ||||||||||||
Db         25 GlyPheLysArgPheSerCysLeuSerLeuGlnSerSerTrpAspTyr------------ 40

Qy        132 AGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGG 191
                 |||   ||||||||||||||    ||||||:::         ||||||||||
Db         41 ---ArgHisAlaProProCysProAlaGly-PheCysLeu------CysValGluThrGl 57

Qy        192 TTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTA 251
              ||||||||||||||||||||                  |||   :::|||
Db         57 yPheHisHisValGlyGlnAlaSer-----------------LeuLysLeuLeuThrSe 71

Qy        252 TTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCA---------CATTTATT 302
              |||   |||         :::   |||         ::::::         |||
Db         71 rSerAspProProAlaLeuAlaSerGlnIleAlaGlyIleThrGlyMetIleHisHisAl 91

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                 ||||||   |||||||||   :::   |||   ||||||||||||   |||
Db         91 aGlnProLeuLeuPheLeuArgGlnGlyValThrLeuLeuProArgLeuGluCysSerSe 111

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
              ||||||   |||||||||   |||   |||   ||||||      ||||||||||||||
Db        111 rAlaIleLeuAlaHisCysIleLeuCysLeuProGlySerSerAspSerProAlaSerAl 131

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
              ||||   ||||||||||||   |||   |||         |||||||||  |||||||||
Db        131 aSerGlnValAlaGlyThrThrGlyValCys---------ProLeuIle--PheValPhe 147

Qy        483 TTAATAAAGACGGGGTTT 500
              :::::::::   ||||||
Db        148 ValValGluMetGlyPhe 153
```

```
RESULT 14
ADE09985
ID   ADE09985 standard; protein; 196 AA.
XX
AC   ADE09985;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Novel protein-related contig polypeptide sequence #573.
XX
KW   novel gene; novel protein; tissue marker; molecular weight marker;
KW   chromosome marker; genetic disorder; contig.
XX
OS   Unidentified.
XX
PN   WO2003054152-A2.
XX
PD   03-JUL-2003.
XX
PF   10-DEC-2002; 2002WO-US039555.
XX
PR   10-DEC-2001; 2001US-0339739P.
PR   11-DEC-2001; 2001US-0339453P.
PR   14-MAR-2002; 2002US-0365091P.
PR   14-MAR-2002; 2002US-0365384P.
PR   12-APR-2002; 2002US-0372381P.
PR   12-APR-2002; 2002US-0372615P.
PR   22-APR-2002; 2002US-00128558.
PR   24-APR-2002; 2002US-0376045P.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Asundi V,  Goodrich RW,  Ren F,  Zhang J,  Zhao QA,  Wang J;
PI   Ghosh M,  Xue AJ,  Wehrman T,  Weng G,  Zhou P,  Drmanac RT,  Wang Z;
PI   Ma Y,  Wang D,  Chen R,  Xu C,  Boyle BJ;
XX
DR   WPI; 2003-569235/53.
XX
PT   New polynucleotides, useful for expressing recombinant proteins for
PT   analysis, characterization or therapeutic use, or as markers for tissues
PT   in which the corresponding protein is preferentially expressed.
XX
PS   Disclosure; SEQ ID NO 3051; 1177pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of novel
CC   proteins. The DNA and protein sequences of the invention are useful as:
CC   markers for tissues in which the corresponding protein is preferentially
CC   expressed; as molecular weight markers on gels; as chromosome markers or
CC   tags; to identify chromosomes or to map related gene positions; and to
CC   compare with endogenous DNA sequences in patients to identify potential
CC   genetic disorders. The present amino acid sequence was used in the
CC   exemplification of the invention.
XX
SQ   Sequence 196 AA;

Alignment Scores:
Pred. No.:              6.39e-33      Length:          196
Score:                  355.00        Matches:         93
Percent Similarity:     62.4%         Conservative:    13
Best Local Similarity:  54.7%         Mismatches:      47
Query Match:            39.4%         Indels:          19
DB:                     7             Gaps:            4

US-10-607-806-1_COPY_5000_5500 (1-501) x ADE09985 (1-196)

Qy         12 TCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCT 71
              ||||||||||||||||||||:::|||   ||||||   |||||||||   |||   |||
Db         10 SerValAlaGlnAlaGlyValGlnTrpCysAspLeuGlySerLeuGlnAlaProProPro 29

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
              ||||||      ||||||||||||||||||   ||||||||||||
Db.        30 GlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAspTyr------------ 45
```

```
Qy        132 AGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAG-AGACAG 190
Db         46 ---ArgArgProProProArgProAlaAsnPheLeu------TyrPhe******ArgArg 62

Qy        191 GTTTTCACCATGTTGGCCAGGCTG-----------------GGTGACCTCCCTTTTAGA 232
Db         63 GlyPheThrValLeuAlaArgMetValSerIleSer***ProArgAspProProAlaSer 82

Qy        233 TTCTCCTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTAT 292
Db         83 AlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg-LeuGlnIleIlePh 102

Qy        293 CACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGG 352
Db        102 eLeuPhe---------PhePhePheSerGluThrGluSerArgSerValAlaGlnAlaGl 119

Qy        353 AGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTC 412
Db        119 yValGlnTrpArgAspLeuGlySerArgHisAlaProProProGlyPheThrProPheSe 139

Qy        413 CCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTT 472
Db        139 rCysLeuSerPheProSerSerTrpAspTyrArgGlyProProProArgProAlaAsnPh 159

Qy        473 TTT-TGTATTTTTAATAAAGACGGGGTT 499
Db        159 eCysCysIlePheSerArgAspGlyVal 168
```

RESULT 15
AAU33304
ID AAU33304 standard; protein; 296 AA.
XX
AC AAU33304;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #3795.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 760; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 296 AA;

Alignment Scores:

Pred. No.:	1.25e-32	Length:	296
Score:	353.00	Matches:	98
Percent Similarity:	47.8%	Conservative:	13
Best Local Similarity:	42.2%	Mismatches:	47
Query Match:	39.2%	Indels:	75
DB:	4	Gaps:	5

US-10-607-806-1_COPY_5000_5500 (1-501) x AAU33304 (1-296)

```
Qy         13 CTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTG 72
Db         56 LeuSerProArgLeuGluCysAsnGlyValIleSerAlaHisCysAsnLeuHisProPro 75

Qy         73 GGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTA 132
Db         76 GlySerSerAspSerProAlaSerAlaAlaArgValAlaGly--------------Ile 90

Qy        133 GCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTT---TTTTTGGTAGAGACA 189
Db         91 ThrGlyThrCysHisHisAlaGlnLeuIlePhePhePheValPheLeuValGluThr 110

Qy        190 GGTTTTCACCATGTTGGCCAGGCTGGG------------------------------ 216
Db        111 GlyPheHisHisAlaAlaGlnAlaGlySerGlnThrProAspLeuArg***SerThrPro 130

Qy        217 ----------------------------------------TGACCTCCCTTTTAGATT 234
Db        131 LeuGlyPheProLysCys***AspTyrArgArgAlaAlaIleValProGlyIlePheLeu 150

Qy        235 CTC---------------------------------------------------------- 237
Db        151 LeuHis***IleArg***LysValProThrLeuLeuThrAspMetArgAsnAlaSerGlu 170

Qy        237 ---------------------------------------------------------- 237
Db        171 TyrAspCysAspPheSerThrAsnLysIleAspLysGluGluThrPheSer***AsnAla 190

Qy        238 ---------CTCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCT--- 285
Db        191 SerLeuAsnLeuCysLeuLeuSerThr***ProTyrGluMetValThrHisPheLysGly 210

Qy        286 TACTTATCACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTT-GTCGCC 344
Db        211 Tyr***IleLeuProLeuPhePhePhePheLeuArgGlnSerLeuSerSerValThr 230

Qy        345 AAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACG 404
Db        231 GlnAlaGlyValGlnTrpHisAsnLeuGlySerLeuGlnProLeuProProGlyPheLys 250

Qy        405 CCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCC 464
Db        251 GlnPheSerCysLeuSerLeuProSerSerTrpAspTyrArgTyrGlnLeuProArgLeu 270

Qy        465 GCTAATTTTTTTGTATTTTTAATAAAGACGGGGTTT 500
Db        271 AlaAsnPhePheValPheLeuValGluThrGlyPhe 282
```

Search completed: February 8, 2006, 09:59:47
Job time : 86.692 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:43:51 ; Search time 3.68824 Seconds
(without alignments)
2613.961 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acggggctcactctgtcgcc.....tttaataaagacggggtttc 501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091821_26876/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_92_@runat_08022006_091821_26876 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	360.5	40.1	627	4	A40201	artifact-warning s
	2	341	37.9	613	4	C40201	artifact-warning s
c	3	322	35.8	613	4	C40201	artifact-warning s
c	4	305.5	33.9	627	4	A40201	artifact-warning s
	5	230	25.6	673	4	F40201	artifact-warning s
c	6	227	25.2	597	4	E40201	artifact-warning s
c	7	194.5	21.6	673	4	F40201	artifact-warning s
	8	187.5	20.8	579	4	D40201	artifact-warning s
	9	185	20.6	100	2	A46010	X-linked retinopat
	10	180	20.0	499	2	S65657	alpha-1C-adrenergi
	11	178.5	19.8	597	4	E40201	artifact-warning s
c	12	168.5	18.7	196	2	I38022	hypothetical prote
	13	156	17.3	841	1	I78885	serine/threonine-s
c	14	136	15.1	579	4	D40201	artifact-warning s
	15	128	14.2	407	2	T02670	probable thromboxa
	16	126	14.0	369	2	A53959	thromboxane A-2 re
	17	123	13.7	79	2	A56194	thromboxane A-2 re
c	18	119.5	13.3	418	2	S41044	chromosomal protei
	19	119	13.2	46	2	I54375	gene NF2 protein -
	20	118	13.1	39	2	I54374	gene NF2 protein -
	21	112	12.4	53	2	A42442	integrin beta-1 ch
	22	100	11.1	301	4	B40201	artifact-warning s
	23	92	10.2	440	2	A26359	decay-accelerating
	24	89	9.9	252	2	T48681	hypothetical prote
	25	86	9.6	806	2	F72024	cell division prot
	26	86	9.6	806	2	F86600	cell division prot
c	27	85	9.4	357	2	E85049	probable transposo
c	28	83	9.2	301	4	B40201	artifact-warning s
	29	83	9.2	522	2	T08711	gamma-adaptin homo
	30	82.5	9.2	320	1	A05023	cytochrome c-type
	31	81	9.0	347	2	T01452	translation initia
c	32	75.5	8.4	394	2	E90115	26S protease regul
c	33	75.5	8.4	816	1	A59003	phosphoinositide 3
c	34	74.5	8.3	404	2	G82449	amino acid biosynt
c	35	74.5	8.3	601	2	JH0170	octopamine recepto
	36	74	8.2	285	2	JC5639	1-acylglycerol-3-p
	37	74	8.2	731	2	D36808	hypothetical prote
	38	73	8.1	373	1	DERTH1	3beta-hydroxy-Delt
	39	73	8.1	373	1	DERTH2	3beta-hydroxy-Delt
	40	73	8.1	373	1	DERTHM	3beta-hydroxy-Delt
c	41	73	8.1	571	2	B86150	hypothetical prote
c	42	73	8.1	679	2	S02165	regulatory protein
c	43	72.5	8.1	738	2	A87516	dipeptidyl peptida
	44	72.5	8.1	962	2	T05845	hypothetical prote
c	45	72.5	8.1	1417	2	F96613	hypothetical prote

ALIGNMENTS

RESULT 1
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:

Pred. No.:	1.03e-30	Length:	627
Score:	360.50	Matches:	94
Percent Similarity:	56.5%	Conservative:	10
Best Local Similarity:	51.1%	Mismatches:	40
Query Match:	40.1%	Indels:	40
DB:	4	Gaps:	5

US-10-607-806-1_COPY_5000_5500 (1-501) x A40201 (1-627)

```
Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::   |||||||||:::|||||||||:::|||||||||||   |||||||||   |||
Db        329 SerArgSerValAlaLysAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaPro 348

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125
```

```
Db        349 ProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAsp--------- 365

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
                  :::     |||     |||     :::     |||   |||
Db        366 ------------TyrArgArgProHisTyrAlaArgLeuIlePheCysIlePheSerArg 381

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCC---------CTTTTAGATTCT 236
              |||     ||||||     |||||||||     |||         ::::::
Db        382 AspGlyValSerProTrp***ProGlyTrpSerArgSerProAspLeuValIleArgPro 401

Qy        237 CCTCATCCT-----------------------------------------GCTCTATTC 254
              |||   |||                                          |||
Db        402 ProTrpProProLysValLeuGlyLeuGlnAla***AspArgAlaArg*********Phe 421

Qy        255 TTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATT 314
              |||    |||   |||       ||||||
Db        422 Phe-IlePhe---CysPheTyrPhePhePro---------------------------- 430

Qy        315 ATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGC 374
                  |||:::||||||||||||||||||||||||||||||   ||||||||||||||||
Db        431 -------ArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSerAl 448

Qy        375 TCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGC 434
              |||||||||||||   |||   |||||||||||||||||||||||||||||||   |||||
Db        448 aHisCysLysLeuArgLeuProGlySerArgHisSerProAlaSerAlaSerGlnValAl 468

Qy        435 TGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACG 494
              |||||||   ||||||    |||||| |||     |||||||||||||:::::|||
Db        468 aGlyThrThrGlyAla-ArgThrThr-ProGly***PhePheValPheLeuValGluThr 487

Qy        495 GGGTTT 500
              ||||||
Db        488 GlyPhe 489

RESULT 2
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              1.35e-28        Length:         613
Score:                  341.00          Matches:        91
Percent Similarity:     57.0%           Conservative:   11
Best Local Similarity:  50.8%           Mismatches:     47
Query Match:            37.9%           Indels:         30
DB:                     4               Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x C40201 (1-613)

Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::|||||||||   |||||||||||:::|||||||||||   |||||||||   |||
Db        319 SerHisSerValThrGlnAlaGlyValGlnTrpArgAspLeuGlySerGlnLeuAlaPro 338

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125

Db        339 ProProGlyPheMetProPheSerCysLeuSerLeuLeuArgThrTrpAspTyr-ArgAr 358

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
                          ||||||||||||||||||||  |||:::   |||   |||
Db        358 gPro---------------HisHisAlaGlnLeuIlePhe--CysIle---PheSerArg 371

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCT 245
              :::     ||||||||||||||||||      :::                  ||||||
Db        372 AsnGlyValLeuProCysTrpProGlyTrpSerArgThrProGlyLeuMetIleHisPro 391

Qy        246 GCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTT-------------- 290
                |||                        |||     :::|||
Db        392 ProLeu-----------------------ProProLysValLeuGlyLeuGlnAla*** 403

Qy        291 --------------ATCACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGC 335
                           |||       |||              ||||||||||||||
Db        404 AlaThrAlaProArgPhePhePhe-PhePhePhe*********LeuArgGlnSerLeuTh 423

Qy        336 TTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCT 395
              ||||||||||||||||   :::|||   ||||||||||||||||||||||||||||
Db        423 rLeuSerProArgLeuGluCysAsnGlyGluIleSerAlaHisCysLysLeuHisLeuPr 443

Qy        396 GGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCA 455
              |||     ||||||||||||||||      |||   ||||||   ||||||    |
Db        443 oGlyLeuCysHisSerProAlaSerAlaPheCysValArgGlyThrThrGlyAla-ArgT 463

Qy        456 CCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGGTTT 500
              ||    |||:::   ||||||||||||||:::::   ||||||
Db        463 hr-MetProSer***PhePheValPheLeuValGluMetGlyPhe 477

RESULT 3
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:
Pred. No.:              1.57e-26        Length:         613
Score:                  322.00          Matches:        88
Percent Similarity:     57.8%           Conservative:   16
Best Local Similarity:  48.9%           Mismatches:     39
Query Match:            35.8%           Indels:         38
DB:                     4               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x C40201 (1-613)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              ||||||:::     |||||||| ||||||     |||:::||| :::|||||||:::|
Db        138 AsnProIleSerThrLysAsnThr-LysAsn***LeuGlyMetVal-ArgAlaProValV 157

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              |||||   |||   :::|||||||||||||||||||   :::|||   ||||||||||||
Db        157 alProArgThrGlnLysAlaGluAlaGlyGluTrpHisLysProGlyArgTrpSerLeuG 177

Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
```

```
Db        177 ln***AlaGluIleSerProLeuHisSerSerLeuGlyAspArgValArgLeuCysLeuL 197

Qy        319 ATAATAATAAG----------------------------------------------AATAATA 302
                  |||                                                 |||   :
Db        197 ysLysLysLysLysLys*********ArgGlyAlaValAlaHisAlaCysAsnProS 217

Qy        301 ATAAATGTGATAAGTAAGGAAACTGGATACTGCATTAGAAAGGGGAAGAATAGAGCAGGA 242
                  ::                |||   ||||||                        |
Db        217 erThrLeuGlyGlyLysGlyGlyTrpIle-----------------------------M 227

Qy        241 TGAGGAGAATCTAAAAGGGAGGTCACCCAGCCTGGCCAACATGGTGAAAA-CCTGTCTCT 183
                  |||||          |||:::      |||||||||||||||||:::   |||
Db        227 etArg-ProGlyValArgAsp------GlnProGlyGlnHisGlyLysThrProPheLeu 244

Qy        182 ACCAAAAAAAAACACACAAAAAATTAGCTGGGCATGGTGGTGTGTGCCAGCTACTTGAGAG 123
                ||| |||||||||||||||||   ||||||   :::|||
Db        245 LeuLys------IleGlnLysIleSerTrpAla---TrpCys-GlyArgLeu-------- 258

Qy        122 CTACAAGTCCCAGCTACTTGAGAGGCTGAGGCAAGAGAATTATTTGAACCCAGGAGATGG 63
                  |||   :::      |||||||||||||||||   :::||||||||   ||
Db        259 ----***SerHisValArgArgArgLeuArgGlnGluAsnGlyIleAsnProGlyGlyGl 277

Qy         62 AGGTTGTAGTGATCTGAGATCACGCCACCGCACTCCAGCCTGGGCGACAGAG 11
                |   ||||||:::   ||||||||||||   ||||||||||||   ||||||
Db        277 yAlaCysSerGluProArgSerArgHisCysThrProAlaTrpValThrGlu 294

RESULT 4
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                9.73e-25        Length:         627
Score:                    305.50          Matches:        90
Percent Similarity:       54.3%           Conservative:   12
Best Local Similarity:    47.9%           Mismatches:     38
Query Match:              33.9%           Indels:         48
DB:                       4               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x A40201 (1-627)

Qy        501 GAAACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTT 442
                |||||||||||||||||||||||||||||| |||||||||   |||          ||||||
Db         34 GluThrProSerLeuLeuLysIleGln-LysIleSerArgAla-***CysGlyArgLeu* 53

Qy        441 AGTCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTT 382
                  |||||||||||||||||||||||:::|||||||||||||||||   |||||||||
Db         53 **SerGlnLeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaC 73

Qy        381 GCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCT 322
                |||||||||||||||   |||||||||||||||||||||||:::::|||||||||
Db         73 ysSerGluProArgSerArgHisCysThrProAlaLeuAlaThrGluArgAspSerValS 93

Qy        321 CAATAATAATAAGAATAATAATAAATGTGATAAGTAAGGAAACTGGATACTGCATTAGAA 262
```

```
Db         93 er-----------------------------------------------------GluL 95

Qy        261 AGGGGAAGAATAGA----------------GCAGGATGAGGA--------------G 235
                ||   ||||||:::                ||||||   |||
Db         95 ysAsnLysAsnLysLys***Lys********AlaGlyArgGlyLeuThrProValIleP 115

Qy        234 AATCTAAAAGGGAGGTCACCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAA 175
                  :::      |||            |||       |||:::   :::
Db        115 roAlaLeuTrpGluAlaLysAlaGlyGlySerArgGlyGlnGluIleGluThrIleLeuA 135

Qy        174 AAACA-------------------------CACAAAAAATTAGCTGGGCAT---GGTG 145
                  |||                      :::|||||||||||||||        |
Db        135 laThrThrValLysProArgLeuTyr***LysTyrLysLysLeuAlaGlyArgSerAlaG 155

Qy        144 GTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTTGAGAGGCTGAGGCAAGAGA 85
                ||   ||| |||                ||||||||||   |||   |||:::|||
Db        155 lyAlaCys-Ser----------------ProSerTyrLeuGlyGly***GlyArgArg 168

Qy         84 ATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGATCACGCCACCGCACTCCAG 25
                :::      |||:::|||   |||:::   |||       |||   ||||||||||||||
Db        169 MetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuGln 188

Qy         24 CCTGGGCGACAGAGTGAG 7
                |||   ||||||||||||
Db        189 ProTrpArgGlnSerGlu 194

RESULT 5
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentia
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:
Pred. No.:                1.55e-16        Length:         673
Score:                    230.00          Matches:        79
Percent Similarity:       41.2%           Conservative:   17
Best Local Similarity:    33.9%           Mismatches:     44
Query Match:              25.6%           Indels:         93
DB:                       4               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x F40201 (1-673)

Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
                |||      |||          |||||||||:::|||:::   |||    |||   :::||
Db        358 GlyProCysSerAspThrProAlaGlyValGlnTrpGlnValLeuAla-HisCysSerLe 377

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
                |:::||||||||||:::::|||   ||||||   |||:::|||||||||
Db        377 uAsnLeuLeuGlySerSerAspSerProAlaSerValSerArgValAlaGly-------- 394

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
                    :::   |||         |||:::   ||||||::         ||||||::
Db        395 -------IleThrGlyMetGlyArgHisSerTrpLeuIleTyr------ValPheLeuIl 410

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
```

```
           :||||||   ||||||||||||   ||||||||                    |||
Db        410 eGluThrGlnPheHisHisValAspGlnAlaGly--------------------LeuLy 423

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATT 302
              ||||||   ||||||                    |||      |||       ::
Db        423 sLeuLeuThrSerSer------------------AspLeuProSerTrpSerProLysVa 437

Qy        303 ATTA------------------------------------------------------- 306
              :|||
Db        437 lLeuGly***GlnAla***AlaThrThrProSer*********PheValPheGlyGlyPh 457

Qy        307 -------------------TTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAG 347
                                  |||   ||||||   :::||||||||||:::|||||
Db        457 ePhePhePhePhePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProAr 477

Qy        348 GCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCA 407
              |||||||   |||||| :::      |||||||||:::||||||   ||||||   ||||
Db        477 gLeuGluCysSerGly-LysPheTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaI 497

Qy        408 TTCTC------------------------------------------------------- 412
              |||||
Db        497 leLeuLeuProLeuSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly* 517

Qy        412 ------------------------------------------------------------ 412

Db        517 **PheMetTyrPhe******ArgHisSerPheThrMetLeuIleArgLeuValSerAsnS 537

Qy        413 -------------------CCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTG 452
                                |||||    ||||||      ||||||   ||||||
Db        537 er***ProGlnValIleCysProPro-GlyLeuProLysCysTrpAspAspArgArgGlu 556

Qy        453 CCACCACGCCCCGCTAATTTTTTTGTATTTTTA 485
              ||||||   ||||||         :::||||||
Db        557 ProProHisProAla*********LeuPheLeu 567

RESULT 6
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                3.27e-16        Length:         597
Score:                    227.00          Matches:        73
Percent Similarity:       50.6%           Conservative:   14
Best Local Similarity:    42.4%           Mismatches:     56
Query Match:              25.2%           Indels:         29
DB:                       4               Gaps:           10

US-10-607-806-1_COPY_5000_5500 (1-501) x E40201 (1-597)

Qy        500 AAACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTA 441
              ||||||   ||||||   ||||||   |||||||||   |||||||||:::   |||:::
Db         36 LysProHisLeuTyr***LysTyr---LysIleSerTrpAlaTrpTrpArgMetProVal 54

Qy        440 GTCCCAGCTACTGGGGAGGCTGAGGCGGGAGAA---TGGCGTGAACCCAGCAGGTGGAGC 384
```

```
              :::|||||||||   |||||||||   ||||||              |||         |||
Db         55 IleProAlaThrArgGluAlaGluThrGlyGluAsnCysLeuAsnProGly------Ser 72

Qy        383 TTGCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGT 324
                          ||||||   |||   ||||||   ||||||:::|||::::::
Db         73 LysValCysGlyGluIleValProSerTyrSerGlyLeuGlyAsnLysSerLysThrPro 92

Qy        323 CTCAATAATAATAAGAATAATAAT---AAATGTGATAAGTAAGGAAACTGGATACTGCAT 267
                       |||            :::|||         |||:::      |||
Db         93 SerGlnLysLysLys*********AlaArgCys---------GlySerSerCysLeu*** 109

Qy        266 TAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGGGAGGT------------ 219
                 :::             ||||||               :::||||||
Db        110 SerGlnArgPheGlySerSerArg--------------ArgGlyGlySerProGluVal 124

Qy        218 ---------CACCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACA 168
                      ||||||   |||||||||      |||:::||||| |||
Db        125 GlySerLeuArgProAla***ProThrTrpArgAsnProIleSerThr-LysAsnIle-- 143

Qy        167 CAAAAAATTAGCTGGGCATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCT 108
              ||||||||||||   ||||||   |||                          ||||||||
Db        144 ----LysLeuAlaGlyArgGlyGlyAlaCys--------------Leu***SerGlnLe 157

Qy        107 ACTTGAGAGGCTGAGGCAAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCT 48
              ||||   ||||||||||||||                  :::      :::   ::
Db        157 uLeuGlyArgLeuArgGlnGluArgIleAla***ThrGlnGluAlaLysPheAla---Va 176

Qy         47 GAGATCACGCCACCGCACTCCAGCCTGGGCGACA 14
              :|||      ||||||||||||||||||||||||
Db        176 lArgLeuCysHisArgThrProAlaTrpAlaThr 187

RESULT 7
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                1.11e-12        Length:         673
Score:                    194.50          Matches:        71
Percent Similarity:       46.0%           Conservative:   16
Best Local Similarity:    37.6%           Mismatches:     57
Query Match:              21.6%           Indels:         45
DB:                       4               Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x F40201 (1-673)

Qy        501 GAAACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTT 442
              ||||||   ||||||||||||||     |||||||||      ||||||      |||
Db         35 GluThrValSerLeuLeuLysIleHis--LysLeuAlaArgAsnGlyGlyProCysLeu* 54

Qy        441 AGTCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTT 382
              |||||||||   ||||||   |||:::||||||   :::||||||   |||      |
Db         54 **SerGlnLeuHisGlyArgGlnArgGlnGluAsnArgLeuAsnProGlyGly***GlyC 74

Qy        381 GCAGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCT 322
```

```
Db         74 ysSerGluProLys-LeuAlaThrAlaLeuGlnProGlyCysGlnSerLysGlyLeuSer 93

Qy        321 CAATAATAATAAGAATAATAATAAATGTGATAAGTAAGGAAACTGGATACTGCATTAGAA 262
Db         94 Gln--------------------------------LysGlnLysGlnSerLysLysLys 102

Qy        261 AGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGGG---------------------- 223
Db        103 LysLysLysThrProLysAsnLys********AlaGlyCysGlyGlySerArgLeuSer 122

Qy        222 ------------------------------------------------AGGTCACCCAGC 211
Db        123 SerGlnHisPheGlyArgProGlyGlyGlnIleThr***GlyGlnGluPheGluThrSer 142

Qy        210 CTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACACAAAAAATTAGCTGGGC 151
Db        143 LeuIleAsnMetValLysLeuCysLeuTyr***Lys------TyrIleAsn***ProGly 160

Qy        150 ATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTTGAGAGGCTGAGGC 91
Db        161 MetAlaAla---------------HisAlaCysAsnProSerTyrThrGlyAspArgGly 175

Qy         90 AAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGATCACGCCACCGCA 31
Db        176 ArgArgIleAla***ThrGlnGluValGluAlaAlaVal-SerGlnAsnLeuProLeuHi 195

Qy         30 CTCCAGCCTGGGCGACAGA 12
Db        195 sSerSerArgGlyValArg 201
```

RESULT 8
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	6.38e-12	Length:	579
Score:	187.50	Matches:	74
Percent Similarity:	50.6%	Conservative:	16
Best Local Similarity:	41.6%	Mismatches:	57
Query Match:	20.8%	Indels:	31
DB:	4	Gaps:	5

US-10-607-806-1_COPY_5000_5500 (1-501) x D40201 (1-579)

```
Qy          3 GGGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCT 62
Db        392 GlySerTrpTyrValAlaGlnAlaGlyValGlnArgCysAsnHisGlySerLeuSerLeu 411

Qy         63 CCATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAG 122
Db        412 AspPheProGly--GlnValIleLeuLeuProGlnProPheGlu***LeuGlyLeu---- 429

Qy        123 CTCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGT 182
Db        430 ----------GlnAlaTyrAlaThrArg-SerGly***PheLeuTyrLeuPhePhe-Va 445

Qy        183 AGAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCAT 242
Db        445 lGluValGlyPheArgHisValAlaGlnAla----------------ValLeuGluLe 459

Qy        243 CCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTA----------- 291
Db        459 uLeuSerTrp----------------SerIleHisLeuProTrpLeuLeuLysCysTr 473

Qy        292 ----------TCACATTTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTC 341
Db        473 pAspTyrArgArgAlaAlaMetLeuGly*********PheArgGlnGlyLeuGlyMetLe 493

Qy        342 GCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTC 401
Db        493 uProArgLeuGluTyrArgGlyAlaIleMetAlaHis***AlaLeu-ThrSerGlnValL 513

Qy        402 ACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGC 461
Db        513 ys***SerSerCysPro-SerLeuLeuSerSerTrpAspTyrArgLeuMetLeuProAsp 532

Qy        462 CCCGCTAATTTTTTTGTA---TTTTTAATAAAGACGGGGTTT 500
Db        533 LeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPhe 546
```

RESULT 9
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked r
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:

Pred. No.:	1.1e-11	Length:	100
Score:	185.00	Matches:	42
Percent Similarity:	77.0%	Conservative:	5
Best Local Similarity:	68.9%	Mismatches:	11
Query Match:	20.6%	Indels:	3
DB:	2	Gaps:	0

US-10-607-806-1_COPY_5000_5500 (1-501) x A46010 (1-100)

```
Qy        321 GAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTG 380
Db          7 GluThrGluSerCysSerValAlaGluAlaGlyValGlnTrpCysAspLeuGlySerLeu 26

Qy        381 CAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGAC 440
Db         27 LysSerProProPro-GlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIl 46

Qy        441 TAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGG 497
Db         46 eThrGlyMetHisHisHisThrGlnLeuIle--PheValPheLeuValGluThrGly 64
```

RESULT 10
S65657
alpha-1C-adrenergic receptor splice form 2 - human
N;Alternate names: alpha-1C-adrenoceptor isoform 2
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 09-Jul-2004
C;Accession: S65657; S65655

```
R;Tanaka, T.
submitted to the EMBL Data Library, July 1994
A;Reference number: S65656
A;Accession: S65657
A;Molecule type: mRNA
A;Residues: 1-499 <TAN>
A;Cross-references: UNIPROT:Q13675; UNIPARC:UPI000004E75E; EMBL:D32202; NID:g927208; PID
R;Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.;
FEBS Lett. 363, 256-260, 1995
A;Title: Cloning, functional expression and tissue distribution of human alpha(1C)-adren
A;Reference number: S65654; MUID:95255557; PMID:7737411
A;Accession: S65655
A;Molecule type: mRNA
A;Residues: 424-499 <HIR>
A;Cross-references: UNIPARC:UPI00001778AB; EMBL:D32202
C;Genetics:
A;Gene: GDB:ADRA1C; ADRA1L1
A;Cross-references: GDB:128088; OMIM:104221
A;Map position: 8p21-8p11.2
C;Superfamily: vertebrate rhodopsin
C;Keywords: alternative splicing; neurotransmitter receptor

Alignment Scores:
Pred. No.:              4.14e-11        Length:         499
Score:                  180.00          Matches:        40
Percent Similarity:     66.7%           Conservative:   2
Best Local Similarity:  63.5%           Mismatches:     19
Query Match:            20.0%           Indels:         2
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_5000_5500 (1-501) x S65657 (1-499)

Qy         22 AGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTGGGTTCAAAT 81
              ||||||||||||   |||:::|||      |||   ||||||   |||   ||||||
Db        431 ArgLeuGluCysSerGlyMetIleLeuAlaHisCysAsnLeuArgLeuProGlySerArg 450

Qy         82 AATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGCTGGCACA 141
              :::|||   ||||||||||||||   |||||||||                    |||
Db        451 AspSerProAlaSerAlaSerGlnAlaAlaGlyThrThrGlyAspValProProGlyArg 470

Qy        142 CACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTTTCACCAT 201
                 |||   ||||||||||||||||        ||||||||||||||||||||||||||
Db        471 ArgHisGlnAlaGlnLeuIlePhe------ValPheLeuValGluThrGlyPheHisHis 488

Qy        202 GTTGGCCAG 210
              |||||||||
Db        489 ValGlyGln 491

RESULT 11
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              6.07e-11        Length:         597
Score:                  178.50          Matches:        62
Percent Similarity:     46.2%           Conservative:   16
Best Local Similarity:  36.7%           Mismatches:     62
Query Match:            19.8%           Indels:         29
DB:                     4               Gaps:           9

US-10-607-806-1_COPY_5000_5500 (1-501) x E40201 (1-597)

Qy         31 TGCGGTGGC---------GTGATCTCAGATCACTACAACCTCCATCTCCTGGGTTCAAAT 81
              |||   |||         :::       |||      |||      |||||||||:::
Db        412 CysProGlyArgSerThrMetAlaGlnSerHisArgLysLeuCysPheLeuGlySerSer 431

Qy         82 AATTCT---CTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGCTGGC 138
              ||||||   |||::: ||||||   :::   |||||||||   |||
Db        432 AsnSerLeuLeuSer-GlnProProGlu***LeuGlyLeuGlnAlaCys----------- 447

Qy        139 ACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTTTCAC 198
                 |||   ||||||   |||::::::      |||    ||||||      ||||
Db        448 ----AlaThrThrProSer***PheTyrIle------Phe-SerArgAspGlyValSerP 464

Qy        199 CATGTTGGCCAGGCTGG------GTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTAT 252
              ||||||||      |||           :::|||      |||      |||
Db        464 roCysTrpSerSerTrpSerGlnThrProAsnLeuArg***SerAlaProPro***Thr- 483

Qy        253 TCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTA 312
              |||      :::
Db        484 -----ProLysAlaLeuGlyLeuGlnAla***AlaThrThrProGly*********PheP 502

Qy        313 TTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCG 372
                  :::                |||||:::||||||||||:::|||   :::|||
Db        502 hePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValArgTrpHisAsnLeu- 521

Qy        373 GCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTC--------------TCCCGCC 417
                                  |||   |||      :::                |||   |
Db        522 -----------------ThrAlaAsnPheAlaSerTrpValGlnAlaIleLeuSerCysL 536

Qy        418 TCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTG 477
              |||||||||||||||||||||   |||      |||||||||||||||||   |||:
Db        536 euSerLeuProSerSerTrpAspTyrArgHisAlaProProArgProAlaAsn---PheI 555

Qy        478 TATTTTTAATAAAGACGGGGTTT 500
              ::||||||::::::   ||||||
Db        555 lePheLeuValGluMetGlyPhe 562

RESULT 12
I38022
hypothetical protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C;Accession: I38022
R;Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L
Cancer Res. 50, 5658-5667, 1990
A;Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A;Reference number: I38021
A;Accession: I38022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI

Alignment Scores:
Pred. No.:              7.05e-10        Length:         196
Score:                  168.50          Matches:        51
Percent Similarity:     56.0%           Conservative:   14
Best Local Similarity:  44.0%           Mismatches:     38
Query Match:            18.7%           Indels:         13
DB:                     2               Gaps:           3

US-10-607-806-1_COPY_5000_5500 (1-501) x I38022 (1-196)

Qy        495 CCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAGTCCC 436
              |||||||||||||||||||  :::|||||||||   |||||||||   |||   ||||
Db         19 ProSerLeuLeuLysIleGln--GluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerG 38
```

```
Qy        435 AGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGCAGTG 376
Db         38 lnLeuLeuArgArgLeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerG 58

Qy        375 AGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCAATAA 316
Db         58 luProLysSerHisLeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysG 78

Qy        315 TAATAAGAATA------------ATAATAAATGTGATAAGTAAGGAAACTGGATACTGCA 268
Db         78 lnAsnLysThrLysAsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCysI 98

Qy        267 TTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGG---------GAGGTCA 217
Db         98 leIleGlyGlyProGlu------------GlyGluGluLysGluTrpSerThrGluMetA 114

Qy        216 CCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAA 173
Db        114 rgSerGluGluLeuMetThrAspAsnValSerIleLeuLysLys 128
```

RESULT 13
I78885
serine/threonine-specific protein kinase (EC 2.7.1.-) STK2 - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Dec-2004
C;Accession: I78885
R;Levedakou, E.N.; He, M.; Baptist, E.W.; Craven, R.J.; Cance, W.G.; Welcsh, P.L.; Simmc
Oncogene 9, 1977-1988, 1994
A;Title: Two novel human serine/threonine kinases with homologies to the cell cycle regu
A;Reference number: I58396; MUID:94268838; PMID:8208544
A;Accession: I78885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-841 <RES>
A;Cross-references: UNIPROT:P51957; UNIPARC:UPI000012FF29; GB:L20321; NID:g348244; PIDN:
C;Genetics:
A;Gene: GDB:STK2
A;Cross-references: GDB:374125
A;Map position: 3p21.1-3p21.1
C;Superfamily: protein kinase homology
C;Keywords: phosphotransferase
F;4-261/Domain: protein kinase homology <KIN>

Alignment Scores:

Pred. No.:	1.71e-08	Length:	841
Score:	156.00	Matches:	33
Percent Similarity:	73.9%	Conservative:	1
Best Local Similarity:	71.7%	Mismatches:	12
Query Match:	17.3%	Indels:	0
DB:	1	Gaps:	0

US-10-607-806-1_COPY_5000_5500 (1-501) x I78885 (1-841)

```
Qy        325 CAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAAG 384
Db        456 GlnSerLeuAlaLeuSerProLysLeuGluCysSerGlyThrIleLeuAlaHisSerAsn 475

Qy        385 CTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAAA 444
Db        476 LeuArgLeuLeuGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThr 495

Qy        445 GGCGCCTGCCACCACGCC 462
Db        496 GlyValCysHisHisAla 501
```

RESULT 14
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t
.

Alignment Scores:

Pred. No.:	2.51e-06	Length:	579
Score:	136.00	Matches:	60
Percent Similarity:	45.8%	Conservative:	21
Best Local Similarity:	33.9%	Mismatches:	54
Query Match:	15.1%	Indels:	42
DB:	4	Gaps:	4

US-10-607-806-1_COPY_5000_5500 (1-501) x D40201 (1-579)

```
Qy        493 GTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAGTCCCAG 434
Db        134 LeuTyrLysLysLysIleGlnLysLeuAlaArgSerGlySerIleSerLeu***SerGln 153

Qy        433 CTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGCAGTGAG 374
Db        154 LeuLeuLysArgLeuGlyGlnGluAspHisLeu-ThrTrpGluValLys-AlaGln***A 173

Qy        373 CCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTC------- 321
Db        173 laMetIleAlaProLeuTyrSerSerLeuGlyAsnIleProArgProCysLeuLys**** 193

Qy        320 --------------------------------AATAATAATAAGAATAATAATAAATGTG 293
Db        193 *****ProSerMetAlaAlaArgLeu***SerGlnHisPheArgSerHisGlyArgCys- 212

Qy        292 ATAAGTAAGGAAACTGGATACTGCATTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAA 233
Db        213 ----------------------------------------------------MetLeuGlnL 216

Qy        232 TCTAAAAGGGAGGTCACCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAA 173
Db        216 eu---ArgSerSerArgThrAlaTrpAlaThrCysArgAsnProThrSerThr--LysLy 234

Qy        172 ACACACAAAAAATTAGCTGGGCATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCC 113
Db        234 sArgTyrLysAsn***ProAspLeuValAla---------------***AlaCysSerPr 249

Qy        112 CAGCTACTTGAGAGGCTGAGGCAAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGT 53
Db        249 oSerTyrSerLysGlyTrpGlyArgArgIle--Thr***ProGlyLysSerArgLeuSer 268

Qy         52 GATCTGAGATCACGCCACCGCACTCCAGCCTGGGCGACA 14
Db        269 GluPro***LeuHisLeuCysThrProAlaTrpAlaThr 281
```

RESULT 15
T02670
probable thromboxane A2 receptor isoform beta - human
C;Species: Homo sapiens (man)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: T02670
R;Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhart-Schultz, K.;
J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; A
 P.; Quan, G.; Kronmiller, B.; Arellano, A.; Montgomery, M.; Ow, D.; Nolan, M.
submitted to the EMBL Data Library, June 1998
A;Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.

A;Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine
A;Reference number: Z14696
A;Accession: T02670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-407 <LAM>
A;Cross-references: UNIPROT:O75228; UNIPARC:UPI00000711C4; EMBL:AC005175; NID:g3253116;
C;Superfamily: prostaglandin E receptor EP1

Alignment Scores:

Pred. No.:	1.83e-05	Length:	407
Score:	128.00	Matches:	30
Percent Similarity:	60.4%	Conservative:	2
Best Local Similarity:	56.6%	Mismatches:	21
Query Match:	14.2%	Indels:	0
DB:	2	Gaps:	0

US-10-607-806-1_COPY_5000_5500 (1-501) x T02670 (1-407)

```
Qy        322 AGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGC 381
              |||:::||||||   |||   |||   ||||||||||||||   ||||||||||||||
Db        328 ArgArgSerLeuThrLeuTrpProSerLeuGluTyrSerGlyThrIleSerAlaHisCys 347

Qy        382 AAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACT 441
                 |||   |||   ||||||      |||   ||||||||||||      ||||||
Db        348 AsnLeuArgLeuProGlySerSerAspSerArgAlaSerAlaSerArgAlaAlaGlyIle 367

Qy        442 AAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTAT 480
                 |||      |||   |||            |||:::
Db        368 ThrGlyValSerHisCysAlaArgProCysMetLeuPhe 380
```

Search completed: February 8, 2006, 10:03:35
Job time : 24.4412 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 24.3157 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_5000_5500
Perfect score: 900
Sequence: 1 acggggctcactctgtcgcc.....tttaataaagacgggggtttc 501

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091819_26843/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_694_@runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	391	43.4	170	2	Q8NAI3_HUMAN	Q8nai3 homo sapien
2	386	42.9	585	1	ALU5_HUMAN	P39192 homo sapien
3	366	40.7	587	1	ALU2_HUMAN	P39189 homo sapien
4	363.5	40.4	591	1	ALU8_HUMAN	P39195 homo sapien
c 5	362	40.2	587	1	ALU2_HUMAN	P39189 homo sapien
6	359	39.9	593	1	ALU7_HUMAN	P39194 homo sapien
7	358.5	39.8	179	2	Q6ZSI5_HUMAN	Q6zsi5 homo sapien
c 8	351.5	39.1	591	1	ALU8_HUMAN	P39195 homo sapien
9	350	38.9	587	1	ALU3_HUMAN	P39190 homo sapien
10	344.5	38.3	158	2	Q8NAL9_HUMAN	Q8nal9 homo sapien
11	342.5	38.1	591	1	ALU1_HUMAN	P39188 homo sapien
c 12	339	37.7	585	1	ALU5_HUMAN	P39192 homo sapien
c 13	337	37.4	593	1	ALU7_HUMAN	P39194 homo sapien
14	333	37.0	593	1	ALU6_HUMAN	P39193 homo sapien
c 15	331	36.8	587	1	ALU3_HUMAN	P39190 homo sapien
c 16	328	36.4	593	1	ALU6_HUMAN	P39193 homo sapien
c 17	323.5	35.9	591	1	ALU1_HUMAN	P39188 homo sapien
18	314.5	34.9	375	2	O60448_HUMAN	O60448 homo sapien
19	312	34.7	603	1	ALU4_HUMAN	P39191 homo sapien
20	311	34.6	239	2	Q6ZWA9_HUMAN	Q6zwa9 homo sapien
21	295.5	32.8	195	2	Q6ZUW2_HUMAN	Q6zuw2 homo sapien
22	293	32.6	200	2	Q6ZUK3_HUMAN	Q6zuk3 homo sapien
23	288	32.0	334	2	Q6ZR97_HUMAN	Q6zr97 homo sapien
24	284.5	31.6	143	2	Q6ZSJ7_HUMAN	Q6zsj7 homo sapien
25	280	31.1	176	2	Q6ZR92_HUMAN	Q6zr92 homo sapien
c 26	277.5	30.8	603	1	ALU4_HUMAN	P39191 homo sapien
27	273	30.3	143	2	Q6ZSB5_HUMAN	Q6zsb5 homo sapien
28	267.5	29.7	241	2	Q6ZTY3_HUMAN	Q6zty3 homo sapien
29	264	29.3	155	2	Q6ZSW8_HUMAN	Q6zsw8 homo sapien
30	262.5	29.2	170	2	Q6ZVI4_HUMAN	Q6zvi4 homo sapien
31	262	29.1	162	2	Q9H5D5_HUMAN	Q9h5d5 homo sapien
32	260.5	28.9	124	2	Q6ZSG7_HUMAN	Q6zsg7 homo sapien
33	260.5	28.9	129	2	Q6ZUW8_HUMAN	Q6zuw8 homo sapien
34	260.5	28.9	176	2	Q6ZTU6_HUMAN	Q6ztu6 homo sapien
35	255.5	28.4	358	2	Q6ZRP1_HUMAN	Q6zrp1 homo sapien
36	250.5	27.8	131	2	Q6ZTD3_HUMAN	Q6ztd3 homo sapien
37	246.5	27.4	350	2	Q96NA4_HUMAN	Q96na4 homo sapien
38	244.5	27.2	160	2	Q6ZN95_HUMAN	Q6zn95 homo sapien
39	243.5	27.1	156	2	Q8NBH4_HUMAN	Q8nbh4 homo sapien
40	243	27.0	134	2	Q6ZSG8_HUMAN	Q6zsg8 homo sapien
41	238.5	26.5	153	2	Q6ZNX9_HUMAN	Q6znx9 homo sapien
c 42	236	26.2	135	2	Q6ZRW7_HUMAN	Q6zrw7 homo sapien
43	235.5	26.2	120	2	Q9H6G8_HUMAN	Q9h6g8 homo sapien
44	234.5	26.1	120	2	Q8N2C8_HUMAN	Q8n2c8 homo sapien
45	234.5	26.1	150	2	Q6ZR91_HUMAN	Q6zr91 homo sapien

ALIGNMENTS

RESULT 1
Q8NAI3_HUMAN
ID Q8NAI3_HUMAN PRELIMINARY; PRT; 170 AA.
AC Q8NAI3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35322.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Prostate;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

```
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK092641; BAC03932.1; -; mRNA.
SQ   SEQUENCE   170 AA;  18511 MW;  310DE58D608976DC CRC64;

Alignment Scores:
Pred. No.:              4.11e-33        Length:         170
Score:                  391.00          Matches:        96
Percent Similarity:     60.8%           Conservative:   14
Best Local Similarity:  53.0%           Mismatches:     34
Query Match:            43.4%           Indels:         37
DB:                     2               Gaps:           5

US-10-607-806-1_COPY_5000_5500 (1-501) x Q8NAI3_HUMAN (1-170)

Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::|||   |||   ||||||||||||||||||||   |||   ||||||||||||||
Db          9 SerHisCysValThrGlnAlaGlyValArgTrpArgGlyLeuSerSerLeuGlnProPro 28

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125
              ||||||||||||:::   ||||||||||||   ||| ||||||:::|||
Db         29 SerProGlyPheGlnArgPheSerCysLeuGlyLeu-GlnValSerGly----------- 44

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
              :::||||||   ||||||||||||||||||||||          ||||||||||
Db         45 ----IleAlaGlyAlaHisHisHisAlaGlnLeuIlePhe------ValPheLeuValGl 61

Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCT 245
              |   |||||||||||||||||||||||||||                  |||   :::||
Db         61 uMetGlyPheHisHisValGlyGlnAlaGly-----------------LeuGluLeuLe 75

Qy        246 GCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCA---------- 294
              |      ::::::|||                  |||   |||:::
Db         75 uThrSerAlaAsnPro-------------------ProSerLeuAlaSerArgGlyAl 88

Qy        295 ------------------------------CATTTATTATTATTCTTATTATTATTGAG 323
                                           |||           :::   |||:::||
Db         88 aGlyIleThrGlyMetSerHisLeuThrGlnHisPhePhePheLeuIlePheLeuMetAr 108

Qy        324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
              ||||||||||:::||||||||||||||||||   ||||||   ||||||||||||||||
Db        108 gGlnSerLeuSerLeuSerProArgLeuGluCysSerGlyThrIleSerAlaHisCysAs 128

Qy        384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
              |||   |||      |||   :::||||||   ||||||      ||||||||||||
Db        128 nLeuCysLeuProCysSerSerAsnSerProThrSerAlaTyrTrpValAlaGlyThrTh 148

Qy        444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTTT 500
              |||      |||||||||  |||     ||||||||||||::::::   ||||||
Db        148 rGlyValHisHisHisAlaGlnLeu--SerPheValPheLeuValGluMetGlyPhe 166

RESULT 2
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.:              1.9e-32         Length:         585
Score:                  386.00          Matches:        99
Percent Similarity:     62.9%           Conservative:   8
Best Local Similarity:  58.2%           Mismatches:     43
```

```
Query Match:          42.9%      Indels:        20
DB:                   1          Gaps:          6

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU5_HUMAN (1-585)

Qy          8 TCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATC 67
              |||||||||||| |||||||||||:::|||||||||||    ||||||||||||||
Db        399 SerLeuCysArg-GlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProPr 418

Qy         68 TCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTC 127
                ||||||||||||   ||||||||||||||||||   ||||||||||||
Db        418 oProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyr-------- 435

Qy        128 AAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGA 187
                          ||||||   ||||||||||||   |||:::   |||   |||||
Db        436 -------ArgArgAlaProProArgProAlaAsnPhe---CysIle---PheSerArgAs 451

Qy        188 CAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTT---------TTAGATTCTCC 238
              |   ||||||||||||||||||||      ||||||          :::       ||
Db        451 pGlyValSerProCysTrpProGlyTrpSerArgSerLeuAspLeuValIleArgProPr 471

Qy        239 TCATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATT 298
              |  |||           |||   ::::::        :::                |||
Db        471 oArgPro-----------ProLysValLeu------GlyLeuGlnAla***AlaThrAl 485

Qy        299 TATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACA 358
                                  ||||||||||||      ||||||  |||||||||   |
Db        485 aProGly*********PhePheGluThrGluSerArgSerValAla-ArgLeuGluCysS 505

Qy        359 GTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCT 418
              ||||||||||||||||||||||   |||   |||   ||||||      ||||||||||
Db        505 erGlyAlaIleSerAlaHisCysAsnLeuArgLeuProGlySerSerAspSerProAlaS 525

Qy        419 CAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGT 478
              ||||||||   ||||||||||||   ||||||   |||||||||   ||||||  |||||
Db        525 erAlaSerArgValAlaGlyThrThrGlyAlaArgHisHisAlaGlnLeuIle--PheVa 544

Qy        479 ATTTTTAATAAAGACGGGGTTT 500
              |||||||::::::|||||||||
Db        544 lPheLeuValGluThrGlyPhe 551

RESULT 3
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION         1     96       Frame-1.
FT   REGION       100    194       Frame-2.
FT   REGION       198    292       Frame-3.
FT   REGION       296    391       Frame-4.
FT   REGION       395    489       Frame-5.
FT   REGION       493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:            2.77e-30      Length:        587
Score:                366.00        Matches:       103
Percent Similarity:   44.1%         Conservative:  12
Best Local Similarity: 39.5%        Mismatches:    42
Query Match:          40.7%         Indels:        105
DB:                   1             Gaps:          8

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU2_HUMAN (1-587)

Qy          7 CTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCAT 66
              |||   ||||||||||||||||||||   |||   ||||||   |||      |||
Db        302 LeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSerAlaHisCysLysLeuArg 321

Qy         67 CTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCT 126
              |||   ||||||   :::|||   |||||||||||:::||||||||||||
Db        322 LeuProGlySerArgHisSerProAlaSerAlaSerArgValAlaGlyThr--------- 338

Qy        127 CAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAG 186
                  |||      ||||||||:::|||||||||           ||||||||||||
Db        339 ------ThrGlyAlaArgHisHisAlaArgLeuIlePhe------ValPheLeuValGlu 354
```

```
Qy        187 ACAGGTTTTCACCA---------------------------------------------- 200
              ||||||||||||
Db        355 ThrGlyPheHisArgValSerGlnAspGlyLeuAspLeuLeuThrSer***SerAlaArg 374

Qy        201 ---------------TGTTGG---------------------------------------- 206
                            ||||||
Db        375 LeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgProAla********* 394

Qy        207 ------------------------------CCAGGCTGG---------------------- 215
                                           |||||||||
Db        395 PhePhe***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArg 414

Qy        216 GTGACCTCC--------------------------CTTTTAGATTCTCCT---------- 239
              :::|||:::                        ||||||      |||
Db        415 LeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProProGlu*** 434

Qy        240 ------------------------CATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTAT 275
                                       |||     |||       |||       :::
Db        435 LeuGlyLeuGlnAlaProAlaThrThrProGly***PheLeuTyrPhe******ArgArg 454

Qy        276 CCAGTTTCCTTACTTATCACATTTATTATTATT--------------------------- 308
                  ::::::|||         :::   |||
Db        455 GlyPheThrValLeuAlaArgMetValSerIleSer***ProArgAspProProAlaSer 474

Qy        309 ----------------------------------------------CTTATTATTATT 320
Db        475 AlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*********PhePhe 494

Qy        321 GAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTG 380
              ||||||||||||      ||||||:::|||||||||||||||   ||||||||||||||
Db        495 GluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeu 514

Qy        381 CAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGAC 440
              ||||||||||||   ||||||||||||||||||   |||||||||||||||||||||||
Db        515 GlnAlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAsp 534

Qy        441 TAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
                  ||||||   ||||||||||||||||||  ||||||||||:::::|||||||||
Db        535 TyrArgArgProProProArgProAlaAsn--PheCysIlePheSerArgAspGlyVal 553
```

```
RESULT 4
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;
```

```
Alignment Scores:
Pred. No.:              5.17e-30        Length:         591
Score:                  363.50          Matches:        102
Percent Similarity:     44.3%           Conservative:   14
Best Local Similarity:  38.9%           Mismatches:     41
Query Match:            40.4%           Indels:         106
DB:                     1               Gaps:           7

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU8_HUMAN (1-591)

Qy          7 CTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCAT 66
              |||   ||||||||||||||||||||   |||   ||||||   |||   ||||||
Db        304 LeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuArg 323

Qy         67 CTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCT 126
              |||   ||||||::::::|||   |||||||||||||:::|||||||||
Db        324 LeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGly------------ 339

Qy        127 CAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAG 186
                  :::   |||          |||||||||:::|||||||||          ||||||||||||
Db        340 ---IleThrGlyAlaArgHisHisAlaArgLeuIlePhe------ValPheLeuValGlu 356
```

```
Qy        187 ACAGGTTTTCACCATGTTGGCCAGGCTGGGTG---------------------------- 218
              ||||||||||||||||||||||||||||||
Db        357 ThrGlyPheHisHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspLeuPro 376

Qy        218 ------------------------------------------------------------ 218

Db        377 ProLysValLeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro********* 396

Qy        218 ------------------------------------------------------------ 218

Db        397 PhePhe***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArg 416

Qy        219 ---------ACCTCC--------------------CTTTTAGATTCTCCT--------- 239
                       ||||||                    ||||||      |||
Db        417 LeuThrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProGlu*** 436

Qy        240 ------------------------CATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTAT 275
                                       |||      |||      |||      :::
Db        437 LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPhe******ArgArg 456

Qy        276 CCAGTTTCCTTACTTATCACATTTATT-----------------ATTATTCTTATT--- 314
              ::::::|||      :::                   :::|||   :::
Db        457 GlyPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIlePheLeuPro 476

Qy        315 ------------------------------------------------------ATT 317

Db        477 LysCysTrpAspTyrArgArgGluProProArgProAlaSerAlaArg********Phe 496

Qy        318 ATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCA 377
              ||||||||||||      ||||||:::|||||||||||||||   ||||||||||||
Db        497 PheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySer 516

Qy        378 CTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGG 437
              ||||||  ||||||  ||||||      ||||||  ||||||||||||||||||||
Db        517 LeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrp 536

Qy        438 GACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGG 497
              |||   ||||||   ||||||||||||||||||  ||||||||||:::::|||||||
Db        537 AspTyrArgArgAlaProProArgProAlaAsn--PheCysIlePheSerArgAspGlyV 556

Qy        498 TT 499
              ||
Db        556 al 556
```

```
RESULT 5
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63742 MW;  3EAAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.:              7.51e-30       Length:         587
Score:                  362.00         Matches:        92
Percent Similarity:     59.5%          Conservative:   11
Best Local Similarity:  53.2%          Mismatches:     41
Query Match:            40.2%          Indels:         29
DB:                     1              Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU2_HUMAN (1-587)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              |||||||||      |||||||||  |||||||||   |||||||||:::||||||:::|
Db        134 AsnProValSerThrLysAsnThr--LysIleSerArgAlaTrpTrpArgAlaProValV 153

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              ||||||||||   ||||||||||||||||||||||||||||   |||   |||||||
Db        153 alProAlaThrArgGluAlaGluAlaGlyGluTrpArgGluProGlyArgArgSerLeuG 173
```

```
Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
Db        173 ln***AlaGluIleAlaProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuL 193

Qy        319 ATAAT----------------------AATAAGAATAATAATAAATGTGATAAGTAAG 284
Db        193 ysLys*********ProGlyAlaValAlaHisAlaCysAsnPro-SerThrLeuGlyGly 212

Qy        283 GAAACTGGATACTGCATTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGG 224
Db        213 ArgGlyGlyArgIleThrArgSerGly------------------------------- 221

Qy        223 GAGGTCACCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACACAAA 164
Db        222 AspArgAspHisProGly***HisGlyGluThrProSerLeuLeuLys------IleGln 239

Qy        163 AAATTAGCTGGGCATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTT 104
Db        240 LysLeuAlaGlyArgGlyGly--------------GlyArgLeu***SerGlnLeuLeu 254

Qy        103 GAGAGGCTGAGGCAAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGA 44
Db        255 GlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArg 274

Qy         43 TCACGCCACCGCACTCCAGCCTGGGCGACAGAG 11
Db        275 SerArgHisCysThrProAlaTrpAlaThrGlu 285
```

RESULT 6
ALU7_HUMAN
ID   ALU7_HUMAN   STANDARD;   PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC   into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC   consensus sequences have been constituted that contain all six
CC   frames conceptual translations of each of these classes of Alu
CC   repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC   codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC   pollution of protein sequence databases with Alu-derived amino
CC   acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC   primate genomes with an average spacing of 4 kb. Some of them are
CC   actively transcribed by pol III. Normal transcripts may contain
CC   Alu-derived sequences in 5' or 3' untranslated regions. However,
CC   cDNA libraries also contain partial and/or rearranged cDNAs
CC   ligated with Alu-derived sequence in any orientation. Although Alu
CC   elements (especially situated on the complementary strand) have a
CC   great potential to create additional/alternative exons,
CC   consideration should be given to the possibility that the presence
CC   of an Alu in an open reading frame may have resulted from a
CC   cloning artifact or may be due to misinterpretation of sequencing
CC   data. This point has been overlooked on several occasions, with
CC   the consequence of erroneous Alu-derived amino acid sequences
CC   being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC   with an Alu-translated entry must be taken as a warning that a
CC   part of Alu repeat may have been artifactually included in the
CC   coding nucleotide sequence.
CC   ---
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION   1   97   Frame-1.
FT   REGION   101   196   Frame-2.
FT   REGION   200   295   Frame-3.
FT   REGION   299   395   Frame-4.
FT   REGION   399   494   Frame-5.
FT   REGION   498   593   Frame-6.
SQ   SEQUENCE   593 AA; 64455 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:   1.59e-29   Length: 593
Score: 359.00   Matches: 91
Percent Similarity: 54.1%   Conservative: 7
Best Local Similarity: 50.3%   Mismatches: 38
Query Match: 39.9%   Indels: 45
DB: 1   Gaps: 5

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU7_HUMAN (1-593)

```
Qy         15 GTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTGGG 74
Db        308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProGly 327

Qy         75 TTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGC 134
Db        328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyr--------------- 342

Qy        135 TGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTT 194
Db        343 ArgArgProProProArgProAlaAsnPhe---CysIle---PheSerArgAspGlyVal 360

Qy        195 TCACCATGTTGGCCAGGCTGGGTGACCTCC---------------------------- 224
Db        361 SerProCysTrpProGlyTrpSerArgThrProAspLeuArg***SerThrArgLeuGly 380

Qy        225 --------------------CTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCCCTTT 263
Db        381 LeuProLysCysTrpAspTyrArgArgGluProProArgProAla*********PhePhe 400

Qy        264 CTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTATTGAG 323
```

```
Db        401 ----------------------------------------------------:::***Asp 402

Qy        324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
                  |||    |||   ||||||||||||   |||||||||||||||||||||||||
Db        403 GlyValSer-LeuLeuLeuProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAs 422

Qy        384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
                |||   |||   ||||||      ||||||||||||||||||   |||||||||
Db        422 nLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleTh 442

Qy        444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTTT 500
                ||||||   |||||||||   ||||||  ||||||||||||::::::|||||||||
Db        442 rGlyAlaArgHisHisAlaArgLeuIle--PheValPheLeuValGluThrGlyPhe 460

RESULT 7
Q6ZSI5_HUMAN
ID   Q6ZSI5_HUMAN PRELIMINARY;      PRT;   179 AA.
AC   Q6ZSI5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein PLJ45499.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Thalamus;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK127408; BAC86963.1; -; mRNA.
SQ   SEQUENCE   179 AA;  19118 MW;  52286CCCA5A4458E CRC64;

Alignment Scores:
Pred. No.:              1.37e-29        Length:         179
Score:                  358.50          Matches:        92
Percent Similarity:     61.3%           Conservative:   11
Best Local Similarity:  54.8%           Mismatches:     43
Query Match:            39.8%           Indels:         22
DB:                     2               Gaps:           5

US-10-607-806-1_COPY_5000_5500 (1-501) x Q6ZSI5_HUMAN (1-179)

Qy         12 TCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCT 71
              ||||||:::|||||||||||||:::|||   ||||||   ||||||||||||||||
Db          5 SerValSerGlnAlaGlyValGlnTrpLeuAspLeuSerSerLeuGlnProProProSer 24

Qy         72 GGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGT 131
              |||||||||   ||||||||||||||||||   ||||||||||||
Db         25 GlyPheLysArgPheSerCysLeuSerLeuGlnSerSerTrpAspTyr------------ 40

Qy        132 AGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGG 191
                  |||   ||||||||||||||||    ||||||:::          ||||||||||
Db         41 ---ArgHisAlaProProCysProAlaGly-PheCysLeu------CysValGluThrGl 57

Qy        192 TTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCTGCTCTA 251
              ||||||||||||||||||||                  |||   :::|||
Db         57 yPheHisHisValGlyGlnAlaSer-----------------LeuLysLeuLeuThrSe 71

Qy        252 TTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCA---------CATTTATT 302
                |||   |||         :::   |||         ::::::         |||
Db         71 rSerAspProProAlaLeuAlaSerGlnIleAlaGlyIleThrGlyMetIleHisHisAl 91

Qy        303 ATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGG 362
                  ||||||   |||||||||   :::   |||   ||||||||||||   |||
Db         91 aGlnProLeuLeuPheLeuArgGlnGlyValThrLeuLeuProArgLeuGluCysSerSe 111

Qy        363 TGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGC 422
                ||||||   |||||||||   |||   |||   ||||||      ||||||||||||||
Db        111 rAlaIleLeuAlaHisCysIleLeuCysLeuProGlySerSerAspSerProAlaSerAl 131

Qy        423 CTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTT 482
                ||||   ||||||||||||   |||   |||          |||||||| ||||||||
Db        131 aSerGlnValAlaGlyThrThrGlyValCys---------ProLeuIle--PheValPhe 147

Qy        483 TTAATAAAGACGGGGTTT 500
                :::::::::   ||||||
Db        148 ValValGluMetGlyPhe 153

RESULT 8
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
```

```
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  64434 MW;  AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.:              1.03e-28        Length:         591
Score:                  351.50          Matches:        94
Percent Similarity:     62.5%           Conservative:   11
Best Local Similarity:  56.0%           Mismatches:     45
Query Match:            39.1%           Indels:         18
DB:                     1               Gaps:           6

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU8_HUMAN (1-591)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              |||||||||      |||||||||  |||||||||   |||||||||:::||||||::::
Db         36 AsnProValSerThrLysAsnThr--LysIleSerArgAlaTrpTrpArgAlaProValI 55

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              ::|||||||||    ||||||||||||||||||      ||||||   |||      ||||
Db         55 leProAlaThrArgGluAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuG 75

Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
              ||   |||||||||||||||||:::|||||||||||||||:::|||||||||   |||
Db         75 ln***AlaGluIleAlaProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeu- 94

Qy        319 ATAATAATAAGAATAATAATAAATGTGATAAGTAAGGAAACTGG---------ATACTGC 269
                  |||                                |||          ::::::
Db         95 -----LysLys*********AlaGlyGlyGlyArgAlaArgTrpLeuThrProValIleP 113

Qy        268 ATTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGGGAGGTCACCCAGCCT 209
                      ||||||         :::|||          |||:::      ||||||
Db        113 roAlaLeuTrpGluGlu---AspHisLeuArg-SerGlyValArgAsp------GlnPro 129

Qy        208 GGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACACAAAAAATTAGCTGGGCAT 149
              ||||||||||||||     ||||||  |||        :::|||||||||||||
Db        130 GlyGlnHisGlyGluThrProSerLeuLeuLys------IleGlnLysLeuAlaGlyArg 147

Qy        148 GGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTTGAGAGGCTGAGGCAA 89
              ||||||                          ||||||||||||   ||||||||||||
Db        148 GlyGlyAla--------------ArgLeu***SerGlnLeuLeuGlyArgLeuArgGln 162

Qy         88 GAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGATCACGCCACCGCACT 29
              ||||||   ||||||||||||  ||||||||||:::   ||||||||||||   |||
Db        163 GluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysThr 182

Qy         28 CCAGCCTGGGCGACAGAG 11
              ||||||||||||||||||
Db        183 ProAlaTrpAlaThrGlu 188

RESULT 9
ALU3_HUMAN
ID   ALU3_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39190;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB1 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U14569; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63613 MW;  85C4155726DEF235 CRC64;

Alignment Scores:
Pred. No.:              1.49e-28       Length:         587
Score:                  350.00         Matches:        102
Percent Similarity:     43.7%          Conservative:   12
Best Local Similarity:  39.1%          Mismatches:     43
Query Match:            38.9%          Indels:         105
DB:                     1              Gaps:           8

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU3_HUMAN (1-587)

Qy          7 CTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCAT 66
              |||   |||||||||||||||||||||   |||   ||||||   |||       |||
Db        302 LeuAlaLeuSerProArgLeuGluCysSerGlyGlyIleSerAlaHisCysLysLeuArg 321

Qy         67 CTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCT 126
              |||   ||||||   :::|||   |||||||||||||||||||||||||||
Db        322 LeuProGlySerArgHisSerProAlaSerAlaSerGlnValAlaGlyThr--------- 338

Qy        127 CAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAG 186
                  |||      |||:::|||:::|||||||||          ||||||||||||
Db        339 ------ThrGlyAlaArgHisTyrAlaArgLeuIlePhe------ValPheLeuValGlu 354

Qy        187 ACAGGTTTTCACCA---------------------------------------------- 200
              ||||||||||||
Db        355 ThrGlyPheHisArgPheSerArgAspGlyLeuAspLeuLeuThrSer***SerAlaArg 374

Qy        201 ---------------TGTTGG---------------------------------------- 206
                            ||||||
Db        375 LeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArgProAla********* 394

Qy        207 ------------------------------CCAGGCTGG--------------------- 215
                                           |||||||||
Db        395 PhePhe***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaGlySerArg 414

Qy        216 GTGACCTCC----------------------------CTTTTAGATTCTCCT--------- 239
              :::|||:::                          ||||||      |||
Db        415 LeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProProLys*** 434

Qy        240 ------------------------CATCCTGCTCTATTCTTCCCCTTTCTAATGCAGTAT 275
                                       |||      |||      |||      :::
Db        435 LeuGlyLeuGlnAlaProAlaThrThrProGly***PheLeuTyrPhe******ArgArg 454

Qy        276 CCAGTTTCCTTACTTATCACATTTATTATTATT------------------------- 308
                  ::::::|||         :::   |||
Db        455 GlyPheThrValLeuAlaGlyMetValSerIleSer***ProArgAspProProAlaSer 474

Qy        309 ----------------------------------------------CTTATTATTATT 320

Db        475 AlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*********PhePhe 494

Qy        321 GAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTG 380
              ||||||||||||      ||||||:::|||||||||||||||   ||||||||||||||
Db        495 GluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeu 514

Qy        381 CAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGAC 440
              ||||||||||||   ||||||||||||||||||   |||||||||||||||||||||||
Db        515 GlnAlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAsp 534

Qy        441 TAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTT 499
                |||||   |||   ||||||||||||  ||||||||||||:::::|||||||||
Db        535 TyrArgArgProProLeuArgProAlaAsn--PheCysIlePheSerArgAspGlyVal 553

RESULT 10
```

```
Q8NAL9_HUMAN
ID   Q8NAL9_HUMAN PRELIMINARY;      PRT;   158 AA.
AC   Q8NAL9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ35131.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Placenta;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK092450; BAC03893.1; -; mRNA.
DR   Ensembl; ENSG00000198057; Homo sapiens.
SQ   SEQUENCE   158 AA;  17414 MW;  E151503015F2FE34 CRC64;

Alignment Scores:
Pred. No.:              4.36e-28       Length:         158
Score:                  344.50         Matches:        80
Percent Similarity:     53.3%          Conservative:   8
Best Local Similarity:  48.5%          Mismatches:     28
Query Match:            38.3%          Indels:         49
DB:                     2              Gaps:           4

US-10-607-806-1_COPY_5000_5500 (1-501) x Q8NAL9_HUMAN (1-158)

Qy          6 GCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCA 65
              :::|||||||||   ||||||||||:::|||      |||   ||||||||||||
Db          3 SerHisSerValThrGlnAlaGlyValGlnTrpCysGlyLeuGlySerLeuGlnProLeu 22

Qy         66 TCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTC 125
                 ||||||||||||   |||   ||||||||||||   ||||||||||||
Db         23 ProProGlyPheLysArgPhePheCysLeuSerLeuLeuSerSerTrpAspTyr------ 40

Qy        126 TCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGA 185
                                  ||||||      |||||||||   |||:::||||||
Db         41 ---------ArgArgValProProProLeuAlaAsnPhe---CysIlePhePhe------ 54
```

```
Qy        186 GACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATCCT 245

Db         54 ------------------------------------------------------------ 54

Qy        246 GCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATT 305
                  :::   ||||||
Db         55 SerPhePhePhe------------------------------------------------ 58

Qy        306 ATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGC 365
                          |||   ||||||      |||   :::||||||||||||||||
Db         59 --------------GluLysGluSerLeuSerValThrGlnAlaGlyValGlnTrpHis 73

Qy        366 GATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTC 425
              ||||||||||||||||||||   |||   ||||||||||||||||||   |||||||||
Db         74 AspLeuGlySerLeuGlnAlaAlaProProGlyPheThrProPheSerCysLeuSerLeu 93

Qy        426 CCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTA 485
              ||||||||||||:::   ||||||   ||||||   ||||||||||||||||||||||
Db         94 ProSerSerTrpAsnTyrArgArgProProProCysProAlaAsnPhePheValPheLeu 113

Qy        486 ATAAAGACGGGGTTT 500
              ::::::   ||||||
Db        114 ValGluMetGlyPhe 118
```

```
RESULT 11
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;
```

```
Alignment Scores:
Pred. No.:              9.7e-28        Length:          591
Score:                  342.50         Matches:          93
Percent Similarity:     63.5%          Conservative:     15
Best Local Similarity:  54.7%          Mismatches:       46
Query Match:            38.1%          Indels:           16
DB:                     1              Gaps:              5

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU1_HUMAN (1-591)

Qy          4 GGGCTCACTCTGTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTC 63
              ||||||  ||||||||||||||||||||    |||   |||:::   |||   :::|||
Db        303 GlyLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleThrAlaHisCysSerLeu 322

Qy         64 CATCTCCTGGGTTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGC 123
                 |||   ||||||::::::      ||||||||||||:::||||||||||||
Db        323 AspLeuProGlySerSerAspProProAlaSerAlaSerArgValAlaGlyThr------ 340

Qy        124 TCTCAAGTAGCTGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTA 183
                       |||      |||||||||:::|||||||||   |||||||||   |||
Db        341 ---------ThrGlyAlaArgHisHisAlaArgLeuIlePhe---ValPhePhe---Val 355

Qy        184 GAGACAGGTTTTCACCATGTTGGCCAGGCTGGGTGACCTCCCTTTTAGATTCTCCTCATC 243
              ||||||||||||||:::|||   ||||||||                      ||||||
Db        356 GluThrGlyPheHisTyrValAlaGlnAlaGlyLeuGlu------------LeuLeu-Gl 371

Qy        244 CTGCTCTATTCTTCCCCTTTCTAATGCAGTATCCAGTTTCCTTACT------TATCACAT 297
                  |||      |||   |||   :::   |||:::   :::|||         |||
Db        371 ySerSerAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisAr 391

Qy        298 TTATTATTATTCTTATTATTATTGAGACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTAC 357
                            ::::::   |||||||||   |||||||||   |||||||||
Db        391 gAlaArg*********PhePhe***AspArgValSerLeuCysArgProGlyTrpSerAl 411

Qy        358 AGTGGTGCGATCTCGGCTCACTGCAAGCTCCACCTGCTGGGTTCACGCCATTCTCCCGCC 417
              |||   |||||||||||||||||:::||||||                |||||||||   ||
Db        411 aValAlaArgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuPr 431
```

```
Qy        418 TCAGCCTCCCCAGTAGCTGGGACTAAAGGCGCCTGCCACCACGCCCCGCTAATTTTTTTG 477
              ||||||||||:::  |||||||||:::|||  ||||||||||||       ||||||
Db        431 oGlnProProGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly--***PheLeu 450

Qy        478 TATTTTTAATAAAGACGGGGTTTC 501
              ||||||     ||||||||||||
Db        451 TyrPheLeu***ArgArgGlyPhe 458
```

```
RESULT 12
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;
```

```
Alignment Scores:
Pred. No.:              2.31e-27        Length:         585
Score:                  339.00          Matches:        94
Percent Similarity:     61.4%           Conservative:   11
Best Local Similarity:  55.0%           Mismatches:     44
Query Match:            37.7%           Indels:         22
DB:                     1               Gaps:           5

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU5_HUMAN (1-585)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              |||||||||      |||||||||  |||||||||   |||||||||:::||||||:::|
Db        133 AsnProValSerThrLysAsnThr--LysIleSerTrpAlaTrpTrpArgAlaProValV 152

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              ||||||||||   ||||||||||||||||      ||||||   |||       ||||
Db        152 alProAlaThrArgGluAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuG 172

Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
              ||   |||||||||||||||||:::|||||| ||||||||::::::||||||||||||
Db        172 ln***AlaGluIleAlaProLeuHisSerSer-LeuAlaThrGluArgAspSerValSer 191

Qy        319 ---------ATAATAATAAGAATAATAATAAATGTGATAAGTAAGGAAACT---GGATAC 272
                                  :::   :::      :::      |||   |||
Db        192 LysLys*********ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGly--- 210

Qy        271 TGCATTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGGGAGGTCACCCAG 212
                                     |||   |||              |||:::
Db        211 -----------------GlyArgGlyGlyArgIleThrArgSerArgAspArgAspHis 224

Qy        211 CCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACACAAAAAATTAGCTGGG 152
              ||||||||||||||||      ||||||   |||         :::||||||||||||
Db        225 ProGlyGlnHisGlyGluThrProSerLeuLeuLys------IleGlnLysLeuAlaGly 242

Qy        151 CATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTTGAGAGGCTGAGG 92
                ||||||                          ||||||||||   |||||||||
Db        243 ArgGlyGlyAla---------------ArgLeu***SerGlnLeuLeuGlyArgLeuArg 257

Qy         91 CAAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGATCACGCCACCGC 32
              |||||||||   ||||||||||||   ||||||||||||:::   ||||||||||||
Db        258 GlnGluAsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCys 277

Qy         31 ACTCCAGCCTGGGCGACAGAGTGAG 7
              |||||||||||| ||||||||||||
Db        278 ThrProAlaTrp-ArgGlnSerGlu 285
```

```
RESULT 13
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:                3.82e-27        Length:         593
Score:                    337.00          Matches:        91
Percent Similarity:       56.9%           Conservative:   12
Best Local Similarity:    50.3%           Mismatches:     33
Query Match:              37.4%           Indels:         45
DB:                       1               Gaps:           5

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU7_HUMAN (1-593)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              |||||||||       |||||||||  |||||||||   |||||||||:::|||||::::
Db         36 AsnProValSerThrLysAsnThr--LysIleSerArgAlaTrpTrpArgAlaProValI 55

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              ::|||||||||  ||||||||||||||||      ||||||   |||      ||||
Db         55 leProAlaThrArgGluAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgLeuG 75

Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
              ||   |||||||||||||||||:::|||||||||||:::|||::: :::::::   |||
Db         75 ln***AlaGluIleAlaProLeuHisSerSerLeuGlyAsnLysSer-GluThrProSer 94

Qy        319 ATAATAATAAGAATAATAATAAATGTGATAAGTAAGGAAACTGGATACTGCATTAGAAAG 260
                                                                    :::|||
Db         95 ----------------------------------------------------------GlnLys 96

Qy        259 GGGAAGAATAGAGCAGGATGAGGAGAATCTAAA-------------------------- 227
                        ||||||  |||  |||:::
Db         97 Lys*********AlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgPro 116

Qy        226 AGGGAGGTC----------------------ACCCAGCCTGGCCAACATGGTGAAAAC 191
              |||  |||                      ||||||||||||||||||||
Db        117 ArgArgValAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThr 136

Qy        190 CTGTCTCTACCAAAAAAAACACACAAAAAATTAGCTGGGCATGGTGGTGTGTGCCAGCTA 131
                  ||||||  |||         :::||||||||||||  ||||||
Db        137 ProSerLeuLeuLys------IleGlnLysLeuAlaGlyArgGlyGly------------ 150

Qy        130 CTTGAGAGCTACAAGTCCCAGCTACTTGAGAGGCTGAGGCAAGAGAATTATTTGAACCCA 71
                          ||||||||||||  ||||||||||||||||||  |||||||||
Db        151 ---GlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnPro 169

Qy         70 GGAGATGGAGGTTGTAGTGATCTGAGATCACGCCACCGCACTCCAGCCTGGGCGACA 14
              |||  ||||||||||:::  ||||||||||  ||||||||||||||||||
Db        170 GlyGlyGlyGlyCysSerGluProArgSerArgHisCysThrProAlaTrpAlaThr 188

RESULT 14
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
```

```
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              1.04e-26        Length:         593
Score:                  333.00          Matches:        87
Percent Similarity:     52.5%           Conservative:   8
Best Local Similarity:  48.1%           Mismatches:     41
Query Match:            37.0%           Indels:         45
DB:                     1               Gaps:           5

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU6_HUMAN (1-593)

Qy        15 GTCGCCCAGGCTGGAGTGCGGTGGCGTGATCTCAGATCACTACAACCTCCATCTCCTGGG 74
             ||||||||||||||||||:::||||||||||||   |||   |||||||||   ||||||
Db       308 ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProProGly 327

Qy        75 TTCAAATAATTCTCTTGCCTCAGCCTCTCAAGTAGCTGGGACTTGTAGCTCTCAAGTAGC 134
             ||||||   ||||||||||||||||||   ||||||||||||
Db       328 PheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyr--------------- 342

Qy       135 TGGCACACACCACCATGCCCAGCTAATTTTTTGTGTGTTTTTTTTGGTAGAGACAGGTTT 194
                |||   ||||||   ||||||||||||   |||:::   |||   ||||||
Db       343 ArgHisAlaProProArgProAlaAsnPhe---CysIle---PheSerArgAspGlyVal 360

Qy       195 TCACCATGTTGGCCAGGCTGGGTGACCTCC---------------------------- 224
             ||||||||||||   ||||||      :::
Db       361 SerProCysTrpSerGlyTrpSerArgThrProAspLeuArg***SerAlaArgLeuGly 380

Qy       225 --------------------CTTTTAGATTCTCCTCATCCTGCTCTATTCTTCCCCTTT 263
                                  :::   |||   ||||||                |||
Db       381 LeuProLysCysTrpAspTyrArgArgGluProProArgProAla*********PhePhe 400

Qy       264 CTAATGCAGTATCCAGTTTCCTTACTTATCACATTTATTATTATTCTTATTATTATTGAG 323
                                                                     :::
Db       401 ----------------------------------------------------------***Asp 402

Qy       324 ACAGAGTCTTGCTTTGTCGCCAAGGCTGGAGTACAGTGGTGCGATCTCGGCTCACTGCAA 383
                |||    |||   ||||||||||||   :::||||||||||||||||||
Db       403 GlyValSer-LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAs 422

Qy       384 GCTCCACCTGCTGGGTTCACGCCATTCTCCCGCCTCAGCCTCCCCAGTAGCTGGGACTAA 443
              |||   |||   ||||||      ||||||||||||||||||   |||||||||
Db       422 nLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleTh 442

Qy       444 AGGCGCCTGCCACCACGCCCCGCTAATTTTTTTGTATTTTTAATAAAGACGGGGTTT 500
              |||      |||||||||   ||||||  ||||||||||||::::::|||||||||
Db       442 rGlyMetArgHisHisAlaArgLeuIle--PheValPheLeuValGluThrGlyPhe 460

RESULT 15
ALU3_HUMAN
ID   ALU3_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39190;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SB1 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
```

```
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14569; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    194       Frame-2.
FT   REGION      198    292       Frame-3.
FT   REGION      296    391       Frame-4.
FT   REGION      395    489       Frame-5.
FT   REGION      493    587       Frame-6.
SQ   SEQUENCE   587 AA;  63613 MW;  85C4155726DEF235 CRC64;

Alignment Scores:
Pred. No.:              1.7e-26        Length:          587
Score:                  331.00         Matches:         88
Percent Similarity:     57.8%          Conservative:    12
Best Local Similarity:  50.9%          Mismatches:      44
Query Match:            36.8%          Indels:          29
DB:                     1              Gaps:            4

US-10-607-806-1_COPY_5000_5500 (1-501) x ALU3_HUMAN (1-587)

Qy        499 AACCCCGTCTTTATTAAAAATACAAAAAAATTAGCGGGGCGTGGTGGCAGGCGCCTTTAG 440
              |||||||||      |||||||||  |||||||||   |||   |||:::||||||:::|
Db        134 AsnProValSerThrLysAsnThr--LysIleSerArgAla***TrpArgAlaProValV 153

Qy        439 TCCCAGCTACTGGGGAGGCTGAGGCGGGAGAATGGCGTGAACCCAGCAGGTGGAGCTTGC 380
              ||||||||||   ||||||||||||||||||||||||||||   |||   |||||||
Db        153 alProAlaThrTrpGluAlaGluAlaGlyGluTrpArgGluProGlyArgArgSerLeuG 173

Qy        379 AGTGAGCCGAGATCGCACCACTGTACTCCAGCCTTGGCGACAAAGCAAGACTCTGTCTCA 320
              ||   |||||||||   ||||||:::||||||||||||||:::|||||||||   |||
Db        173 ln***AlaGluIleProProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuL 193

Qy        319 ATAAT----------------------AATAAGAATAATAATAAATGTGATAAGTAAG 284
                                       :::      |||    :::   :::
Db        193 ysLys*********ProGlyAlaValAlaHisAlaCysAsnPro-SerThrLeuGlyGly 212

Qy        283 GAAACTGGATACTGCATTAGAAAGGGGAAGAATAGAGCAGGATGAGGAGAATCTAAAAGG 224
              |||         |||   |||
Db        213 ArgGlyGlyArgIleThrArgSerGly-------------------------------- 221

Qy        223 GAGGTCACCCAGCCTGGCCAACATGGTGAAAACCTGTCTCTACCAAAAAAAACACACAAA 164
              :::         ||||||   :::||||||      ||||||   |||          :::
Db        222 AspArgAspHisProGly***AsnGlyGluThrProSerLeuLeuLys------IleGln 239

Qy        163 AAATTAGCTGGGCATGGTGGTGTGTGCCAGCTACTTGAGAGCTACAAGTCCCAGCTACTT 104
              ||||||||||||      |||                                ||||||||||||
Db        240 LysLeuAlaGlyArgSerGly--------------GlyArgLeu***SerGlnLeuLeu 254

Qy        103 GAGAGGCTGAGGCAAGAGAATTATTTGAACCCAGGAGATGGAGGTTGTAGTGATCTGAGA 44
              ||||||||||||||||||   :::|||||||||   |||   ||||||:::   |||
Db        255 GlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArg 274

Qy         43 TCACGCCACCGCACTCCAGCCTGGGCGACAGAG 11
              |||||||||   ||||||||||||||||||||
Db        275 SerArgHisCysThrProAlaTrpAlaThrGlu 285

Search completed: February  8, 2006, 09:43:21
Job time : 127.578 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 16.3058 Seconds
(without alignments)
2694.617 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 928
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091817_26823/app_query.fasta_1
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_605_@runat_08022006_091817_26823 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	350	37.7	110	3	AAB43292	Aab43292 Human ORF
c	2	339.5	36.6	143	5	ADK36945	Adk36945 Novel hum
	3	324	34.9	85	4	AAU20249	Aau20249 Human nov
	4	321	34.6	119	4	AAM83429	Aam83429 Human imm
c	5	318	34.3	260	8	ADR09618	Adr09618 Human pro
c	6	317.5	34.2	132	4	AAM01209	Aam01209 Human pro
c	7	317.5	34.2	132	4	AAU69854	Aau69854 Human pro
c	8	317.5	34.2	132	4	ABU71745	Abu71745 Prostate
c	9	317.5	34.2	132	5	ABB95314	Abb95314 Human P77
c	10	317.5	34.2	132	6	ABR54426	Abr54426 Prostate
c	11	317.5	34.2	132	7	ADB14023	Adb14023 Human pro
c	12	317.5	34.2	132	7	ADG26439	Adg26439 Human pro
c	13	317.5	34.2	135	4	AAM01284	Aam01284 P789P spl
c	14	317.5	34.2	135	4	AAU69929	Aau69929 Human pro
c	15	317.5	34.2	135	4	ABU71820	Abu71820 Prostate
c	16	317.5	34.2	135	5	ABB95389	Abb95389 Human P77
c	17	317.5	34.2	135	6	ABR54501	Abr54501 Prostate
c	18	317.5	34.2	135	7	ADB14334	Adb14334 Human pro
c	19	317.5	34.2	135	7	ADG26750	Adg26750 Human pro
c	20	317	34.2	94	5	ABJ03693	Abj03693 Human ova
c	21	317	34.2	123	7	ADB63958	Adb63958 Human pro
	22	317	34.2	136	6	ADA54195	Ada54195 Human pro
	23	316	34.1	85	4	AAO11283	Aao11283 Human pol
c	24	314	33.8	158	7	ADB64893	Adb64893 Human pro
c	25	311.5	33.6	138	7	ADM05509	Adm05509 Human pro
c	26	311	33.5	102	4	AAO08307	Aao08307 Human pol
c	27	310	33.4	124	8	ADQ65229	Adq65229 Novel hum
c	28	309	33.3	103	4	AAO06180	Aao06180 Human pol
c	29	307.5	33.1	202	8	ADQ66597	Adq66597 Novel hum
c	30	307	33.1	103	4	AAU32477	Aau32477 Novel hum
c	31	305.5	32.9	122	8	ADQ65187	Adq65187 Novel hum
	32	305	32.9	80	5	ABP60188	Abp60188 Human mac
	33	305	32.9	118	7	ADM05260	Adm05260 Human pro
c	34	304	32.8	127	5	AAE20800	Aae20800 Human gen
c	35	304	32.8	127	5	ABG64678	Abg64678 Human alb
c	36	304	32.8	127	8	ADL77945	Adl77945 Albumin f
c	37	302.5	32.6	104	7	ADB65611	Adb65611 Human pro
c	38	302.5	32.6	527	7	ADD27133	Add27133 Human adi
c	39	301.5	32.5	166	8	ADQ66566	Adq66566 Novel hum
	40	301	32.4	118	6	ADA54860	Ada54860 Human pro
c	41	300.5	32.4	217	7	ADC87151	Adc87151 Human GPC
	42	299	32.2	81	4	AAU29944	Aau29944 Novel hum
	43	298	32.1	102	7	ADM06071	Adm06071 Human pro
c	44	298	32.1	134	5	AAO21703	Aao21703 Human sec
c	45	298	32.1	136	7	ADB65293	Adb65293 Human pro

ALIGNMENTS

RESULT 1
AAB43292
ID AAB43292 standard; protein; 110 AA.
XX
AC AAB43292;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX

```
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000; 2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC77501.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 11; Page 5297-5298; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ   Sequence 110 AA;

Alignment Scores:
Pred. No.:                4.87e-33        Length:          110
Score:                    350.00          Matches:         63
Percent Similarity:       73.6%           Conservative:    4
Best Local Similarity:    69.2%           Mismatches:      24
Query Match:              37.7%           Indels:          0
DB:                       3               Gaps:            0

US-10-607-806-1_COPY_1_500 (1-500) x AAB43292 (1-110)

Qy        347 CTTTTTTTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCC 288
              |||      ||||||||||||   |||   :::||||||||||||||||||||   |||
Db         16 LeuValAlaGlnAlaGlyVal***TrpHisSerLeuGlySerLeuGlnProProLeuPro 35

Qy        287 AGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCA 228
              ||||||   |||   |||   ||||||      ||||||||||||||||      |||
Db         36 GlyPheLysGlnPheSerCysArgSerLeuProSerSerTrpAspTyrArgHisAlaPro 55

Qy        227 CCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGC 168
              |||      ||||||||||||||||||   ||||||   |||   ||||||||||||||||
Db         56 ProArgGlnAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGly 75

Qy        167 TGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAA 108
              ||||||||||||||   |||:::   ||||||   |||   :::|||||||||||||:::
Db         76 TrpSerGlnThrProAspLeuArgArgSerThrHisLeuSerValProLysCysTrpAsp 95

Qy        107 TACAGACATGAGCCACCGCACCTGGCTAATGAG 75
              ||||||   ||||||||||||||||||   |||
Db         96 TyrArgArgGluProProHisLeuAlaTyrGlu 106

RESULT 2
ADK36945
ID   ADK36945 standard; protein; 143 AA.
XX
AC   ADK36945;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human polypeptide SeqID9027.
XX
KW   antiarthritic; antiparkinsonian; neuroprotective; nootropic;
KW   immunosuppressive; cytostatic; antipsoriatic; antiinflammatory;
KW   antibacterial; antiviral; antifungal; antiparasitic; gene therapy;
KW   arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;
KW   psoriasis; inflammatory bowel disease; infection; bacteria; virus;
KW   fungus; parasite; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .143
FT                   /label= OTHER
FT                   /note= "OTHER= All Xaa's in this sequence are unknown
FT                   amino acids or the site of a stop codon within the DNA
FT                   sequence"
XX
PN   WO200216439-A2.
XX
PD   28-FEB-2002.
XX
PF   05-MAR-2001; 2001WO-US004941.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2002-280918/32.
XX
PT   Isolated polynucleotide encoding bone marrow derived polypeptides useful
PT   for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's
PT   disease, and inflammatory bowel disease.
XX
PS   Claim 20; SEQ ID NO 9027; 504pp; English.
XX
CC   This invention relates to a novel isolated polynucleotide comprising a
CC   nucleotide sequence selected from one of 1680 sequences, a mature protein
CC   coding portion of them, an active domain of them and their complementary
CC   sequences. The invention may be useful for the production of compounds
CC   with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,
CC   immunosuppressive, cytostatic, antipsoriatic, antiinflammatory,
CC   antibacterial, antiviral, antifungal or antiparasitic activity. In
CC   addition, the disclosed sequences may be useful for gene therapy. The
CC   polypeptides or their antibodies are useful for treating many diseases
CC   such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,
CC   psoriasis, inflammatory bowel disease and infections caused by bacteria,
CC   viruses, fungi or parasites. The present sequence is that of a human
CC   polypeptide of the invention.
XX
SQ   Sequence 143 AA;

Alignment Scores:
Pred. No.:                9.9e-32         Length:          143
Score:                    339.50          Matches:         64
Percent Similarity:       68.7%           Conservative:    4
Best Local Similarity:    64.6%           Mismatches:      24
Query Match:              36.6%           Indels:          7
DB:                       5               Gaps:            1

US-10-607-806-1_COPY_1_500 (1-500) x ADK36945 (1-143)

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              |||   ||||||||||||||||||||   ||||||||||||||   |||   ||||||
```

```
Db         28 GlnGlyGlyValGlnTrpCysAsnLeuSerSerLeuGlnProProLeuProGlyPheLys 47

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
              |||   |||  ||||||      ||||||||||||||   ||||||   |||
Db         48 ***PheSerCysLeuSerLeuProSerSerTrpAspTyrArgCysValPro***HisPro 67

Qy        218 GCTAATTTTTGT--------------------ATTTTTATTAGAGATAGTGTTTTGCCA 180
              |||||||||                      ||||||   ||||||   |||   |||
Db         68 ProAsnPheCysLeuPhePheIlePheTyrPheIlePheSerArgAspGluValSerPro 87

Qy        179 TGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCA 120
              ||||||||||||||||||:::||||||   ||||||   ||||||:::||||||||||||
Db         88 CysTrpProGlyTrpSerArgThrProAspLeuLysTrpSerThrArgLeuGlyLeuPro 107

Qy        119 AAGTGCTGGGAATACAGACATGAGCCACCGCACCTGGCTAATGAGTTACTATTAACT 63
              |||||||||:::|||||||||||||||||              :::   |||||||||
Db        108 LysCysTrpAspTyrArgHisGluProAlaAlaGlnProGluLysThrLeuLeuThr 126

RESULT 3
AAU20249
ID   AAU20249 standard; protein; 85 AA.
XX
AC   AAU20249;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human novel endocrine antigen, SEQ ID No 306.
XX
KW   Human; endocrine antigen; cytostatic; antiinfertility; antidiabetic;
KW   thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW   antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW   reproductive disorder; endocrine cancer; pancreatic disorder;
KW   diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW   hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200155319-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001335.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
```

```
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-457726/49.
DR   N-PSDB; AAS32352.
XX
PT   Isolated polypeptide for treating, preventing and prognosing disorders
PT   related to the endocrine system including endocrine disorders,
PT   reproductive disorders, and gastrointestinal disorders and also for
PT   testing and detection e.g. diagnosis.
XX
PS   Claim 11; SEQ ID NO 306; 558pp; English.
XX
CC   The invention relates to cDNAs encoding novel human endocrine antigens or
CC   a fragment having biological activity, a domain, an epitope, full length
CC   protein, variant, allelic variant or a species homologue of the
CC   cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition when administered (e.g. by
CC   gene therapy or antisense-therapy). Identifying mutations in the genes
CC   coding for the antigens is useful for diagnosing a pathological condition
CC   or a susceptibility to a pathological condition. The DNAs, antigens and
CC   antibodies raised against the antigens useful for treating, preventing
CC   and/ or prognosing disorders related to the endocrine system or hormone
CC   imbalance or reproductive disorders, cancers of endocrine tissues,
CC   disorders of the pancreas (e.g. diabetes mellitus), the adrenal glands
CC   (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC   hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC   of diseases and disorders are given in the specification. The present
CC   sequence represents an endocrine antigen of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 85 AA;

Alignment Scores:
Pred. No.:              6.39e-30        Length:         85
Score:                  324.00          Matches:        60
Percent Similarity:     84.4%           Conservative:   5
Best Local Similarity:  77.9%           Mismatches:     12
Query Match:            34.9%           Indels:         0
DB:                     4               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x AAU20249 (1-85)

Qy        109 TCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAG 168
              ||||||||||||||   |||       |||||||||:::|||:::||||||   ||||||
Db          1 SerGlnHisPheGlySerProMetGlnValAspHisMetArgSerGlyValArgAspGln 20

Qy        169 CCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGT 228
              ||||||||||||||||||||   ||||||::::::||||||||||||||||||||||||
Db         21 ProGlyGlnHisGlyLysThrProSerLeuLeuArgIleGlnLysLeuAlaArgCysGly 40

Qy        229 GGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTG 288
              |||   ||||||   |||||||||   |||||||||   |||:::|||   |||||||||
Db         41 GlyAlaCysLeuTrpSerGlnLeuLeuArgArgLeuArgGlnGluAsnArgLeuAsnLeu 60

Qy        289 GGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGG 339
              ||||||||||||||||||||||||||      |||||||||||||||||
Db         61 GlyGlyGlyGlyCysSerGluProArgSerCysHisCysThrProAlaTrp 77

RESULT 4
AAM83429
ID   AAM83429 standard; protein; 119 AA.
XX
AC   AAM83429;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen SEQ ID NO:11022.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
```

PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR N-PSDB; AAK56210.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Claim 11; SEQ ID NO 11022; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 119 AA;

Alignment Scores:

Pred. No.:	1.64e-29	Length:	119
Score:	321.00	Matches:	63
Percent Similarity:	74.2%	Conservative:	6
Best Local Similarity:	67.7%	Mismatches:	23
Query Match:	34.6%	Indels:	1

```
DB:                 4           Gaps:         0

US-10-607-806-1_COPY_1_500 (1-500) x AAM83429 (1-119)

Qy         71 GTAACTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGA 130
              :::|||      ||||||:::||| |||      |||   |||   |||||||||   ||
Db         17 IleThrCysArgProGlySerVal-LeuThrProValIleProThrLeuTrpGluAlaGl 36

Qy        131 GTTGGGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACAC 190
              |   |||||||||   |||   ||||||   ||||||||||||||||||||||   |||
Db         36 uAlaGlyGlySerProGluValArgSerSerArgProAlaTrpProThrTrpGlyAsnLe 56

Qy        191 TATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGC 250
                 |||   |||||||||||||||||:::   ||||||   ||||||::::::|||||
Db         56 uPheSerThrLysAsnThrLysIleSerArgAlaTrpTrpGlnLeuProValIleProAl 76

Qy        251 TACACAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGC 310
              ||||||||||||   |||   ||||||||||||||   ||||||||||||||   ||
Db         76 aThrGlnGluAlaGluAlaGlyGluSerLeuGluProArgArgTrpArgLeuGlnTrpAl 96

Qy        311 CAAGATTGCACCACTGCACTCCAGCCTGGAAAAAAAG 347
              ||||:::|||||||||||||||||||||      |||
Db         96 aLysValAlaProLeuHisSerSerLeuGlySerLys 108

RESULT 5
ADR09618
ID   ADR09618 standard; protein; 260 AA.
XX
AC   ADR09618;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Human protein useful for treating neurological disease Seq 3124.
XX
KW   human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   N-PSDB; ADR07662.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 3124; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polypeptide is a protein
CC   encoded by a full length human cDNA sequence of the invention. NOTE: This
CC   sequence is not given in the sequence listing of the specification but
CC   can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC   office.
XX
SQ   Sequence 260 AA;

Alignment Scores:
Pred. No.:                4.81e-29      Length:          260
Score:                    318.00        Matches:         61
Percent Similarity:       63.1%         Conservative:    4
Best Local Similarity:    59.2%         Mismatches:      22
Query Match:              34.3%         Indels:          16
DB:                       8             Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x ADR09618 (1-260)

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              |||||||||   ||||||||||||||   ||||||||||||||||||||      |||
Db         36 GlnAlaGlyAlaGlnTrpCysAsnLeuSerSerLeuGlnProProProProGluLeuLys 55

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||||||||   |||||||||   ||||||||||||||      ||||||
Db         56 GlyPhePheCysLeuSerLeuLeuThrSerTrpAspTyrArgHisAlaProProArgPro 75

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||||||||   ||||||   |||            ||||||||||:::
Db         76 AlaAsnPheCysIlePheSerArgAspGlyValSer---------ProGlyTrpSerArg 92

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              |||      |||:::   ||||||   |||   ||||||||||||||:::||||||
Db         93 ThrSerAspLeuArgTrpSerThrHisLeuCysLeuProLysCysTrpAspTyrArgArg 112

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTAACTGAGCGCTAACTCCATCCCAGGTCC 39
              ||||||                                        |||||||||   :::
Db        113 GluPro----------------------------------------LeuHisProAlaThr 119

Qy         38 CATATGCTA 30
              |||   |||
Db        120 HisAlaLeu 122

RESULT 6
AAM01209
ID   AAM01209 standard; protein; 132 AA.
XX
AC   AAM01209;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific P776P splice variant contig 6 amino acid #1.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US001574.
XX
PR   14-JAN-2000; 2000US-00483672.
XX
PA   (CORI-) CORIXA CORP.
XX
```

```
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT   monitoring and treating prostate cancer in a patient and for use in
PT   vaccines.
XX
PS   Claim 2; Page 435; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC   prepared using (I) or (II) are used treat cancer in a patient. (I) and
CC   the antibodies are also used in the detection of cancer in a patient. The
CC   cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC   and (II) can be used in vaccines. The antibodies or (I) can be used for
CC   monitoring the progression of cancer in a patient. (I) and (II) can also
CC   be used to improve diagnostic and therapeutic methods for prostate
CC   cancer. They can indicate the level of metastasis as well as the prostate
CC   volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent
CC   polynucleotide and amino acid sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29      Length:          132
Score:                  317.50        Matches:         66
Percent Similarity:     61.1%         Conservative:    14
Best Local Similarity:  50.4%         Mismatches:      38
Query Match:            34.2%         Indels:          13
DB:                     4             Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x AAM01209 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::                 ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||   :::|||   ||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||   |||   ||||||      |||||||||:::|||      ||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||      |||      |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||      ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 7
AAU69854
ID   AAU69854 standard; protein; 132 AA.
XX
AC   AAU69854;
XX
DT   30-JAN-2002  (first entry)
XX
DE   Human prostate cDNA encoded protein #49.
XX
KW   Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US009919.
XX
PR   27-MAR-2000; 2000US-00536857.
PR   09-MAY-2000; 2000US-00568100.
PR   12-MAY-2000; 2000US-00570737.
PR   13-JUN-2000; 2000US-00593793.
PR   27-JUN-2000; 2000US-00605783.
PR   09-AUG-2000; 2000US-00636215.
PR   29-AUG-2000; 2000US-00651236.
PR   06-SEP-2000; 2000US-00657279.
PR   02-OCT-2000; 2000US-00679426.
PR   10-OCT-2000; 2000US-00685166.
PR   09-NOV-2000; 2000US-00709729.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   N-PSDB; AAS63925.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer.
XX
PS   Claim 2; Page 430-431; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific polynucleotides,
CC   polypeptides, fusion proteins of the polypeptides, antibodies raised
CC   against the polypeptides (or antigenic epitopes derived from them) and
CC   antigen-presenting cells expressing the polypeptides. The antibodies are
CC   useful for detecting the presence of cancer, especially prostate cancer.
CC   The polypeptides, polynucleotides and the antigen-presenting cells are
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein, and for inhibiting the development of cancer especially prostate
CC   cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC   useful for stimulating an immune response, and for treating cancer. The
CC   oligonucleotide is useful for detecting cancer. The present sequence is a
CC   prostate specific polypeptide of the invention
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29      Length:          132
Score:                  317.50        Matches:         66
Percent Similarity:     61.1%         Conservative:    14
Best Local Similarity:  50.4%         Mismatches:      38
Query Match:            34.2%         Indels:          13
DB:                     4             Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x AAU69854 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::                 ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39
```

```
Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
              |||   |||   ||||||      |||||||||:::|||      ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||       |||       |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||      ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

```
RESULT 8
ABU71745
ID   ABU71745 standard; protein; 132 AA.
XX
AC   ABU71745;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Prostate cancer specific antigen P776P #1.
XX
KW   Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW   immunogen; cancer; prostate specific antigen.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2002192763-A1.
XX
PD   19-DEC-2002.
XX
PF   29-JUN-2001; 2001US-00895793.
XX
PR   04-OCT-1999;   99US-0157455P.
PR   04-OCT-2000; 2000US-00679272.
PR   28-MAR-2001; 2001US-00822827.
XX
PA   (XUJJ/) XU J.
PA   (DILL/) DILLON D C.
PA   (MITC/) MITCHAM J L.
PA   (HARL/) HARLOCKER S L.
PA   (JIAN/) JIANG Y.
PA   (KALO/) KALOS M D.
PA   (FANG/) FANGER G R.
PA   (RETT/) RETTER M W.
PA   (STOL/) STOLK J A.
PA   (DAYC/) DAY C H.
PA   (VEDV/) VEDVICK T S.
PA   (CART/) CARTER D.
PA   (LISX/) LI S X.
PA   (WANG/) WANG A.
PA   (SKEI/) SKEIKY Y A W.
PA   (HEPL/) HEPLER W T.
PA   (HEND/) HENDERSON R A.
PA   (HURA/) HURAL J.
PA   (MCNE/) MCNEILL P D.
PA   (HOUG/) HOUGHTON R L.
PA   (DBAS/) Y DE BASSOLS C V.
PA   (FOYT/) FOY T M.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA,  Hural J;
PI   Mcneill PD,  Houghton RL,  Y De BassolsCV,  Foy TM;
```

```
XX
DR   WPI; 2001-245062/25.
DR   N-PSDB; ACA59733.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for the
PT   treatment and diagnosis of prostate cancer.
XX
PS   Example 5; SEQ ID NO 573; 85pp; English.
XX
CC   The invention describes a fusion protein comprising at least one amino
CC   acid sequence of immunogenic portions of any of the 3 sequences not
CC   defined in the specification, or sequences having at least 70 or 90 %
CC   sequence identity to any one of the 35 sequences defined in the USPTO web
CC   site, which is encoded by any of the 4 nucleotide sequences not defined
CC   in the specification. The fusion protein, composition and methods are
CC   useful for diagnosing, preventing and/or treating cancer, particularly
CC   prostate cancer. The proteins are useful as markers to indicate the
CC   presence or absence of cancer. This is the amino acid sequence of a
CC   prostate cancer specific antigen. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the US patent office at
CC   seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ   Sequence 132 AA;
```

```
Alignment Scores:
Pred. No.:                4.49e-29        Length:          132
Score:                    317.50          Matches:         66
Percent Similarity:       61.1%           Conservative:    14
Best Local Similarity:    50.4%           Mismatches:      38
Query Match:              34.2%           Indels:          13
DB:                       4               Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x ABU71745 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||       :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC---------------------------- 339
              ::::::   :::                  ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
              |||   |||   ||||||      |||||||||:::|||      ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||       |||       |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||      ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

```
RESULT 9
ABB95314
ID   ABB95314 standard; protein; 132 AA.
XX
AC   ABB95314;
XX
DT   19-JUL-2002  (first entry)
XX
DE   Human P776P splice variant encoded protein SEQ ID NO 573.
XX
```

```
KW   Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2002022248-A1.
XX
PD   21-FEB-2002.
XX
PF   12-JAN-2001; 2001US-00759143.
XX
PR   25-FEB-1997;  97US-00806099.
PR   01-AUG-1997;  97US-00904804.
PR   10-FEB-1998;  98US-00020956.
PR   25-FEB-1998;  98US-00030607.
PR   14-JUL-1998;  98US-00115453.
PR   23-SEP-1998;  98US-00159812.
PR   15-JAN-1999;  99US-00232149.
PR   09-APR-1999;  99US-00288946.
PR   13-JUL-1999;  99US-00352616.
PR   12-NOV-1999;  99US-00439313.
PR   18-NOV-1999;  99US-00443686.
PR   14-JAN-2000; 2000US-00483672.
PR   27-MAR-2000; 2000US-00536857.
PR   09-MAY-2000; 2000US-00568100.
PR   12-MAY-2000; 2000US-00570737.
PR   13-JUN-2000; 2000US-00593793.
PR   27-JUN-2000; 2000US-00605783.
PR   10-AUG-2000; 2000US-00636215.
PR   29-AUG-2000; 2000US-00651236.
PR   06-SEP-2000; 2000US-00657279.
PR   02-OCT-2000; 2000US-00679426.
PR   10-OCT-2000; 2000US-00685166.
XX
PA   (XUJJ/) XU J.
PA   (DILL/) DILLON D C.
PA   (MITC/) MITCHAM J L.
PA   (HARL/) HARLOCKER S L.
PA   (JIAN/) JIANG Y.
PA   (KALO/) KALOS M D.
PA   (FANG/) FANGER G R.
PA   (RETT/) RETTER M W.
PA   (STOL/) STOLK J A.
PA   (DAYC/) DAY C H.
PA   (VEDV/) VEDVICK T S.
PA   (CART/) CARTER D.
PA   (LISX/) LI S X.
PA   (WANG/) WANG A.
PA   (SKEI/) SKEIKY Y A W.
PA   (HEPL/) HEPLER W T.
PA   (HEND/) HENDERSON R A.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2002-255649/30.
XX
PT   New prostate-specific polynucleotides for diagnosing and treating
PT   diseases, in particular prostate cancer, and as markers for the
PT   progression of cancer.
XX
PS   Claim 2; SEQ ID NO 573; 87pp; English.
XX
CC   The present invention provides prostate-specific coding sequences and
CC   their encoded proteins. These can be used in the diagnosis and treatment
CC   of cancers, particularly prostate cancer. The present sequence is a
CC   protein described in the invention
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29        Length:         132
```

```
Score:                  317.50          Matches:        66
Percent Similarity:     61.1%           Conservative:   14
Best Local Similarity:  50.4%           Mismatches:     38
Query Match:            34.2%           Indels:         13
DB:                     5               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x ABB95314 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||       :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::                 ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                |||   |||   ||||||      |||||||||:::|||      ||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||   |||   ||||||   |||      |||      |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||       ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

```
RESULT 10
ABR54426
ID   ABR54426 standard; protein; 132 AA.
XX
AC   ABR54426;
XX
DT   28-AUG-2003  (first entry)
XX
DE   Prostate tumour specific protein sequence SEQ ID 573.
XX
KW   Cytostatic; gene therapy; prostate-specific protein; PSP; human;
KW   immune response; prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO200289747-A2.
XX
PD   14-NOV-2002.
XX
PF   09-MAY-2002; 2002WO-US014753.
XX
PR   09-MAY-2001; 2001US-00852911.
PR   29-JUN-2001; 2001US-00895814.
PR   10-DEC-2001; 2001US-00012896.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Henderson RA;
PI   Kalos MD,  Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS;
PI   Carter D,  Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Hural J;
PI   Mcneill PD,  Houghton RL,  Vinals Y De BassolsC,  Foy TM,  Watanabe Y;
PI   Deng T;
XX
DR   WPI; 2003-167130/16.
XX
PT   New prostate-specific proteins and genes, useful in gene therapy,
PT   particularly for stimulating an immune response in a patient, or treating
```

```
PT   prostate cancer in a patient, as well as for diagnosing prostate cancer
PT   in a patient.
XX
PS   Example 5; Page 497-498; 691pp; English.
XX
CC   The present invention relates to novel prostate-specific proteins (PSP)
CC   and their coding sequences. The PSPs and their coding sequences are
CC   useful for stimulating an immune response in a patient, or for treating
CC   prostate cancer in a patient and for determining, detecting or diagnosing
CC   the presence of a cancer in a patient. The present sequence was used to
CC   illustrate the invention
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29        Length:         132
Score:                  317.50          Matches:        66
Percent Similarity:     61.1%           Conservative:   14
Best Local Similarity:  50.4%           Mismatches:     38
Query Match:            34.2%           Indels:         13
DB:                     6               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x ABR54426 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::               ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||   |||   ||||||      |||||||||:::|||      ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||       |||      |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||       ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 11
ADB14023
ID   ADB14023 standard; protein; 132 AA.
XX
AC   ADB14023;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human prostate specific cDNA P776P splice variant contig 6 protein #1.
XX
KW   Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW   cell therapy; vaccine; T-cell epitope;
KW   class I major histocompatibility complex allele; MHC; prostate cancer;
KW   tumour; antigen presenting cell.
XX
OS   Homo sapiens.
XX
PN   US2003185830-A1.
XX
PD   02-OCT-2003.
XX
PF   12-NOV-2002; 2002US-00294025.
XX
PR   25-FEB-1997;   97US-00806099.
PR   01-AUG-1997;   97US-00904804.
PR   09-FEB-1998;   98US-00020956.
PR   25-FEB-1998;   98US-00030607.
PR   14-JUL-1998;   98US-00115453.
PR   23-SEP-1998;   98US-00159812.
PR   15-JAN-1999;   99US-00232149.
PR   09-APR-1999;   99US-00288946.
PR   13-JUL-1999;   99US-00352616.
PR   12-NOV-1999;   99US-00439313.
PR   18-NOV-1999;   99US-00443686.
PR   14-JAN-2000; 2000US-00483672.
PR   27-MAR-2000; 2000US-00536857.
PR   09-MAY-2000; 2000US-00568100.
PR   12-MAY-2000; 2000US-00570737.
PR   13-JUN-2000; 2000US-00593793.
PR   27-JUN-2000; 2000US-00605783.
PR   09-AUG-2000; 2000US-00636215.
PR   29-AUG-2000; 2000US-00651236.
PR   06-SEP-2000; 2000US-00657279.
PR   02-OCT-2000; 2000US-00679426.
PR   10-OCT-2000; 2000US-00685166.
PR   09-NOV-2000; 2000US-00709729.
PR   12-JAN-2001; 2001US-00759143.
PR   09-FEB-2001; 2001US-00780669.
PR   09-MAY-2001; 2001US-00852911.
PR   29-JUN-2001; 2001US-00895814.
PR   10-DEC-2001; 2001US-00012896.
PR   09-MAY-2002; 2002US-00144678.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Stolk JA,  Kalos MD;
XX
DR   WPI; 2003-756193/71.
DR   N-PSDB; ADB14020.
XX
PT   New isolated polypeptide for use in a vaccine for stimulating an immune
PT   response, or for treating or diagnosis cancer, preferably prostate
PT   cancer.
XX
PS   Example 5; Page; 101pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising no more than
CC   11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
CC   peptides comprise a fragment ADB13563 of that contain naturally processed
CC   T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC   alleles. ADB13563 is a polypeptide encoded by a human prostate specific
CC   cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC   encoding the proteins and peptides, expression vectors, a host cell
CC   transformed with the vector, an isolated antibody (or antigen binding
CC   fragment) that specifically binds to the protein or peptide, detecting
CC   the presence of a cancer in a patient (comprising contacting a patient
CC   sample with a binding agent that binds to the peptides or a polypeptide
CC   appearing as ADB13558, detecting the amount of polypeptide that binds to
CC   the agent and comparing the amount of polypeptide to a predetermined cut-
CC   off value to determine the presence of cancer), a fusion protein
CC   comprising the peptides or proteins, stimulating or expanding T cells
CC   specific for a tumour protein comprising contacting T cells with the
CC   peptides or the isolated T cell population, treating prostate cancer in a
CC   patient comprising administering a composition comprising the peptides,
CC   nucleic acids, antibodies or compounds, determining the presence of a
CC   cancer in a patient and treating prostate cancer in a patient comprising
CC   incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC   from a patient with the peptides or antigen presenting cells that express
CC   (the peptides so that the T cells proliferate, and administering the
CC   proliferated T cells to the patient. The peptides (or an oligonucleotide
CC   that hybridises to nucleic acid encoding them), is used to detect the
CC   presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC   antigen-presenting cells expressing the nucleic acid, are used to
CC   stimulate or expand T cells specific for a tumour protein. The peptides,
```

```
CC   nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC   presenting cells are used to stimulate an immune response or treat
CC   prostate cancer in a patient. The present sequence is a prostate specific
CC   protein of the invention. Note: Except where otherwise indicated, the
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=20030185830.
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29       Length:         132
Score:                  317.50         Matches:        66
Percent Similarity:     61.1%          Conservative:   14
Best Local Similarity:  50.4%          Mismatches:     38
Query Match:            34.2%          Indels:         13
DB:                     7              Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x ADB14023 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::                  ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||   |||   ||||||      |||||||||:::|||    ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||      |||      |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||      ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 12
ADG26439
ID   ADG26439 standard; protein; 132 AA.
XX
AC   ADG26439;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Human prostate-specific polypeptide #95.
XX
KW   Human; prostate-specific polypeptide; prostate cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003157089-A1.
XX
PD   21-AUG-2003.
XX
PF   09-MAY-2002; 2002US-00144678.
XX
PR   25-FEB-1997;   97US-00806099.
PR   01-AUG-1997;   97US-00904804.
PR   09-FEB-1998;   98US-00020956.
PR   25-FEB-1998;   98US-00030607.
PR   14-JUL-1998;   98US-00115453.
PR   23-SEP-1998;   98US-00159812.
PR   15-JAN-1999;   99US-00232149.
PR   09-APR-1999;   99US-00288946.
PR   13-JUL-1999;   99US-00352616.
PR   12-NOV-1999;   99US-00439313.
PR   18-NOV-1999;   99US-00443686.
PR   14-JAN-2000; 2000US-00483672.
PR   27-MAR-2000; 2000US-00536857.
PR   09-MAY-2000; 2000US-00568100.
PR   12-MAY-2000; 2000US-00570737.
PR   13-JUN-2000; 2000US-00593793.
PR   27-JUN-2000; 2000US-00605783.
PR   09-AUG-2000; 2000US-00636215.
PR   29-AUG-2000; 2000US-00651236.
PR   06-SEP-2000; 2000US-00657279.
PR   02-OCT-2000; 2000US-00679426.
PR   10-OCT-2000; 2000US-00685166.
PR   09-NOV-2000; 2000US-00709729.
PR   12-JAN-2001; 2001US-00759143.
PR   09-FEB-2001; 2001US-00780669.
PR   09-MAY-2001; 2001US-00852911.
PR   29-JUN-2001; 2001US-00895814.
PR   10-DEC-2001; 2001US-00012896.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Henderson RA;
PI   Kalos MD,  Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS;
PI   Carter D,  Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Hural J;
PI   Mcneill PD,  Houghton RL,  Vinals Y De BassolsC,  Foy TM,  Watanabe Y;
PI   Meagher MJ,  Deng T;
XX
DR   WPI; 2003-777973/73.
DR   N-PSDB; ADG26436.
XX
PT   New polynucleotides encoding prostate specific polypeptides isolated from
PT   a human prostate tumor cDNA library are useful to diagnose and treat
PT   cancer particularly prostate cancer.
XX
PS   Example 5; SEQ ID NO 573; 99pp; English.
XX
CC   The invention relates to human prostate-specific polypeptides and the
CC   polynucleotides encoding them. The invention also relates to an isolated
CC   antibody or its antigen-binding fragment that specifically binds a
CC   polypeptide of the invention, a method of detecting cancer in a patient
CC   comprising contacting a biological sample of the patient with an agent
CC   that binds a prostate-specific polypeptide and comparing the amount of
CC   bound polypeptide compared to a predetermined cut-off value and a fusion
CC   protein comprising a prostate-specific polypeptide. The sequences of the
CC   invention are used to diagnose and treat cancer, particularly prostate
CC   cancer. This sequence represents a human prostate-specific polypeptide of
CC   the invention. Note: The sequence data for this patent did not form part
CC   of the printed specification but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 132 AA;

Alignment Scores:
Pred. No.:              4.49e-29       Length:         132
Score:                  317.50         Matches:        66
Percent Similarity:     61.1%          Conservative:   14
Best Local Similarity:  50.4%          Mismatches:     38
Query Match:            34.2%          Indels:         13
DB:                     7              Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x ADG26439 (1-132)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::                  ||||||
```

```
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279

Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219

Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159

Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99

Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66

Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

RESULT 13
AAM01284
ID AAM01284 standard; protein; 135 AA.
XX
AC AAM01284;
XX
DT 04-OCT-2001 (first entry)
XX
DE P789P splice variant contig 6 amino acid #2.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US001574.
XX
PR 14-JAN-2000; 2000US-00483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
XX
PS Claim 2; Page 514-515; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 135 AA;

Alignment Scores:

Pred. No.:	4.52e-29	Length:	135
Score:	317.50	Matches:	66
Percent Similarity:	61.1%	Conservative:	14
Best Local Similarity:	50.4%	Mismatches:	38
Query Match:	34.2%	Indels:	13
DB:	4	Gaps:	2

US-10-607-806-1_COPY_1_500 (1-500) x AAM01284 (1-135)

```
Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372

Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339

Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279

Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219

Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159

Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99

Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66

Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

RESULT 14
AAU69929
ID AAU69929 standard; protein; 135 AA.
XX
AC AAU69929;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate cDNA encoded protein #74.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US009919.
XX
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
XX

```
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   N-PSDB; AAS64161.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer.
XX
PS   Claim 2; Page 552; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific polynucleotides,
CC   polypeptides, fusion proteins of the polypeptides, antibodies raised
CC   against the polypeptides (or antigenic epitopes derived from them) and
CC   antigen-presenting cells expressing the polypeptides. The antibodies are
CC   useful for detecting the presence of cancer, especially prostate cancer.
CC   The polypeptides, polynucleotides and the antigen-presenting cells are
CC   useful for stimulating and/or expanding T cells specific for a tumour
CC   protein, and for inhibiting the development of cancer especially prostate
CC   cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC   useful for stimulating an immune response, and for treating cancer. The
CC   oligonucleotide is useful for detecting cancer. The present sequence is a
CC   prostate specific polypeptide of the invention
XX
SQ   Sequence 135 AA;

Alignment Scores:
Pred. No.:              4.52e-29        Length:         135
Score:                  317.50          Matches:        66
Percent Similarity:     61.1%           Conservative:   14
Best Local Similarity:  50.4%           Mismatches:     38
Query Match:            34.2%           Indels:         13
DB:                     4               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x AAU69929 (1-135)

Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg---------- 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::               ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||   |||   ||||||      |||||||||:::|||      ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||      |||      |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||      ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130

RESULT 15
ABU71820
ID   ABU71820 standard; protein; 135 AA.
XX
AC   ABU71820;
XX
DT   10-JUN-2003  (first entry)
XX
DE   Prostate cancer specific antigen P776P #16.
XX
KW   Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW   immunogen; cancer; prostate specific antigen.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US2002192763-A1.
XX
PD   19-DEC-2002.
XX
PF   29-JUN-2001; 2001US-00895793.
XX
PR   04-OCT-1999;   99US-0157455P.
PR   04-OCT-2000; 2000US-00679272.
PR   28-MAR-2001; 2001US-00822827.
XX
PA   (XUJJ/) XU J.
PA   (DILL/) DILLON D C.
PA   (MITC/) MITCHAM J L.
PA   (HARL/) HARLOCKER S L.
PA   (JIAN/) JIANG Y.
PA   (KALO/) KALOS M D.
PA   (FANG/) FANGER G R.
PA   (RETT/) RETTER M W.
PA   (STOL/) STOLK J A.
PA   (DAYC/) DAY C H.
PA   (VEDV/) VEDVICK T S.
PA   (CART/) CARTER D.
PA   (LISX/) LI S X.
PA   (WANG/) WANG A.
PA   (SKEI/) SKEIKY Y A W.
PA   (HEPL/) HEPLER W T.
PA   (HEND/) HENDERSON R A.
PA   (HURA/) HURAL J.
PA   (MCNE/) MCNEILL P D.
PA   (HOUG/) HOUGHTON R L.
PA   (DBAS/) Y DE BASSOLS C V.
PA   (FOYT/) FOY T M.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA,  Hural J;
PI   Mcneill PD,  Houghton RL,  Y De BassolsCV,  Foy TM;
XX
DR   WPI; 2001-245062/25.
DR   N-PSDB; ACA59969.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for the
PT   treatment and diagnosis of prostate cancer.
XX
PS   Example 5; SEQ ID NO 884; 85pp; English.
XX
CC   The invention describes a fusion protein comprising at least one amino
CC   acid sequence of immunogenic portions of any of the 3 sequences not
CC   defined in the specification, or sequences having at least 70 or 90 %
CC   sequence identity to any one of the 35 sequences defined in the USPTO web
CC   site, which is encoded by any of the 4 nucleotide sequences not defined
CC   in the specification. The fusion protein, composition and methods are
CC   useful for diagnosing, preventing and/or treating cancer, particularly
CC   prostate cancer. The proteins are useful as markers to indicate the
CC   presence or absence of cancer. This is the amino acid sequence of a
CC   prostate cancer specific antigen. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the US patent office at
CC   seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ   Sequence 135 AA;

Alignment Scores:
```

Pred. No.:	4.52e-29	Length:	135
Score:	317.50	Matches:	66
Percent Similarity:	61.1%	Conservative:	14
Best Local Similarity:	50.4%	Mismatches:	38
Query Match:	34.2%	Indels:	13
DB:	4	Gaps:	2

US-10-607-806-1_COPY_1_500 (1-500) x ABU71820 (1-135)

```
Qy        431 GGCCTGAAGGATGCTAAGTGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGG 372
              |||      :::|||:::   |||:::   ||||||   ::::::
Db          4 GlyGluGlyGluAlaArgHisValLeuHisGlyGlyArgArgGluArg------------ 19

Qy        371 TTGCAAGTAAAGTTATTAATTACCCTTTTTTTC-------------------------- 339
              ::::::   :::               ||||||
Db         20 ValArgGlyGluThrAlaThrAsnPhePhePheLeuArgGlnGluSerGlyProValAla 39

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::|||   ||||||||||||   |||   |||||||||
Db         40 GlnAlaGlyValGlnTrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLys 59

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                 |||   |||   ||||||      |||||||||:::|||      ||||||||||||
Db         60 GlnPheSerCysLeuSerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeu 79

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||   |||   ||||||   |||       |||       |||||||||:::
Db         80 AlaAsnPheCysSerPheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLys 99

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   |||:::   |||       ||||||||||||||||||||   |||||||||
Db        100 ThrProGlyLeuGlnGlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHis 119

Qy         98 GAGCCACCGCACCTGGCTAATGAGTTACTATTA 66
              :::|||||||||   |||      :::||||||
Db        120 LysProProHisProAlaCysHisIleLeuLeu 130
```

Search completed: February 8, 2006, 09:59:42
Job time : 86.529 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:43:51 ; Search time 3.68088 Seconds
(without alignments)
2613.961 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 928
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091821_26876/app_query.fasta_1
-DB=PIR -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_92_@runat_08022006_091821_26876 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	278	30.0	613	4	C40201	artifact-warning s
2	271	29.2	673	4	F40201	artifact-warning s
c 3	268	28.9	673	4	F40201	artifact-warning s
4	261.5	28.2	613	4	C40201	artifact-warning s
5	239.5	25.8	196	2	I38022	hypothetical prote
c 6	221.5	23.9	627	4	A40201	artifact-warning s
c 7	208.5	22.5	597	4	E40201	artifact-warning s
8	207	22.3	627	4	A40201	artifact-warning s
c 9	188.5	20.3	499	2	S65657	alpha-1C-adrenergi
10	186	20.0	579	4	D40201	artifact-warning s
11	184	19.8	418	2	S41044	chromosomal protei
12	175.5	18.9	597	4	E40201	artifact-warning s
c 13	170.5	18.4	579	4	D40201	artifact-warning s
c 14	166	17.9	100	2	A46010	X-linked retinopat
c 15	158	17.0	301	4	B40201	artifact-warning s
c 16	132	14.2	841	1	I78885	serine/threonine-s
17	126	13.6	301	4	B40201	artifact-warning s
18	116	12.5	114	2	JC5238	galactosylceramide
c 19	115.5	12.4	39	2	I54374	gene NF2 protein -
c 20	115.5	12.4	46	2	I54375	gene NF2 protein -
c 21	115.5	12.4	79	2	A56194	thromboxane A-2 re
c 22	115.5	12.4	407	2	T02670	probable thromboxa
c 23	109	11.7	53	2	A42442	integrin beta-1 ch
24	109	11.7	124	2	A47582	B-cell growth fact
c 25	107	11.5	369	2	A53959	thromboxane A-2 re
26	107	11.5	447	2	A57034	transcription fact
c 27	104	11.2	107	2	C48078	biliary glycoprote
c 28	93.5	10.1	619	2	A60646	transforming prote
c 29	87	9.4	275	1	AG0889	3',5'-cyclic-nucle
c 30	85	9.2	342	2	PC4211	hepatocellular car
c 31	85	9.2	440	2	A26359	decay-accelerating
c 32	84	9.1	252	2	PC4259	ferritin associate
c 33	80.5	8.7	536	2	T41695	hypothetical prote
c 34	78	8.4	522	2	T08711	gamma-adaptin homo
c 35	77	8.3	275	1	F65090	3',5'-cyclic-nucle
c 36	77	8.3	275	1	G85963	3',5'-cyclic-nucle
c 37	77	8.3	275	1	H91118	3',5'-cyclic-nucle
c 38	77	8.3	3947	2	T52486	ferrichrome sidero
39	75	8.1	96	2	S02187	hypothetical prote
40	75	8.1	151	1	B39697	myb-related protei
41	74	8.0	632	2	T48616	hypothetical prote
c 42	74	8.0	764	2	H83055	probable outer mem
c 43	74	8.0	992	2	T46337	hypothetical prote
44	73.5	7.9	253	2	F75403	probable metal bin
c 45	73.5	7.9	333	2	T02614	hypothetical prote

ALIGNMENTS

RESULT 1
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:

Pred. No.:	3.83e-22	Length:	613
Score:	278.00	Matches:	63
Percent Similarity:	59.5%	Conservative:	9
Best Local Similarity:	52.1%	Mismatches:	35
Query Match:	30.0%	Indels:	14
DB:	4	Gaps:	3

US-10-607-806-1_COPY_1_500 (1-500) x C40201 (1-613)

```
Qy        413 TGGGTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGGTTGCAAGTAAAGTTATTA 354
              ||||||           |||      :::|||:::
Db        291 TrpValThrGlu***AspSerValSerLysLysLysLysLysLys*********PhePhe 310

Qy        353 ATTACCCTTTTTTTC----------------------CAGGCTGGAGTGCAGTGGTGC 318
```

```
                    ||||||                              ||||||||||||||||||
Db        311 PhePhePhePhePheGluThrGluSerHisSerValThrGlnAlaGlyValGlnTrpArg 330

Qy        317 AATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTC 258
              :::|||||||||||||||   |||||||||   |||      |||   |||   ||||||
Db        331 AspLeuGlySerLeuGlnAlaProProProGlyPheMetProPheSerCysLeuSerLeu 350

Qy        257 CTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTGGCTAAT------TTTTGTATT 204
              |||   :::|||||||||||||:::      ||||||   |||          |||||||||
Db        351 LeuArgThrTrpAspTyrArgArg------ProHisHisAlaGlnLeuIlePheCysIle 368

Qy        203 TTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAG 144
              |||   |||:::   |||||||||||||||||||||||||||:::||||||   |||  :
Db        369 PheSerArgAsnGlyValLeuProCysTrpProGlyTrpSerArgThrProGlyLeu--M 388

Qy        143 TGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACATGAGCCACCGCACCT 85
              ::|||||||||         ||||||||||||||:::|||      ||||||||||||
Db        388 etIleHisProProLeuProProLysValLeuGlyLeuGlnAla***AlaThrAlaPro 407

RESULT 2
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                2.27e-21        Length:         673
Score:                    271.00          Matches:        58
Percent Similarity:       74.4%           Conservative:   6
Best Local Similarity:    67.4%           Mismatches:     20
Query Match:              29.2%           Indels:         2
DB:                       4               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x F40201 (1-673)

Qy         87 GTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTT 146
              |||   ||||||      |||   ||||||||||||||||||          ||||||||
Db          3 ValTrpTrpLeuThrProValIleProAlaLeuTrpGluThrArgArg-AlaAspHisLe 22

Qy        147 GAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAAT 206
              ||||:::||||||   |||||||||   |||||||||:::|||:::||||||:::|||||
Db         22 uArgSerGlyVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLysIl 42

Qy        207 ACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGG 266
              |   ||||||||||||   ||||||   ||||||   ||||||||||||   |||
Db         42 eHisLysLeuAlaArgAsnGlyGlyProCysLeu***SerGlnLeuHisGlyArgGlnAr 62

Qy        267 GCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTG 326
               |||:::|||   ||||||   ||||||   |||||||||||||:::|||    ||||
Db         62 gGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGluProLysLeu-AlaThrA 82

Qy        327 CACTCCAGCCTGGA 340
              ||||||||||||||
Db         82 laLeuGlnProGly 86

RESULT 3
F40201
artifact-warning sequence (translated ALU class F) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: F40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: F40201
A;Molecule type: DNA
A;Residues: 1-673 <CLA>
A;Cross-references: UNIPARC:UPI000017CECD
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:                4.86e-21        Length:         673
Score:                    268.00          Matches:        57
Percent Similarity:       69.8%           Conservative:   3
Best Local Similarity:    66.3%           Mismatches:     24
Query Match:              28.9%           Indels:         2
DB:                       4               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x F40201 (1-673)

Qy        339 CCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAA 280
              ||||||||||||||||||    :::   ||||||||||||||   ||||||   |||||
Db        589 ProGlyTrpSerAlaVal-AlaSerPheGlySerLeuGlnProGlnProProGlyPheLy 608

Qy        279 GTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCT 220
              |   |||   |||       |||   ||||||||||||||||||      ||||||   ||
Db        608 sArgPheSerCysLeuCysLeuProCysSerTrpAspTyrArgHisGlyProProPheLe 628

Qy        219 GGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCA 160
              |||||||   |||||||||   ||||||:::|||   ||||||      ||||||||||
Db        628 uAlaAsnLeuCysIlePheAsnArgAspThrValSerProCys***SerGlyTrpSerGl 648

Qy        159 AACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACA 100
              |||||||   ||||||   |||      ||| ||||||||||||||||||   |||   :
Db        648 nThrProAspLeuLys***SerAlaLeuLeu-ValSerGlnSerAlaGlyMetThrGlyV 668

Qy         99 TGAGCCACCGCACC 86
              ::||||||   |||
Db        668 alSerHisHisThr 672

RESULT 4
C40201
artifact-warning sequence (translated ALU class C) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: C40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: C40201
A;Molecule type: DNA
A;Residues: 1-613 <CLA>
A;Cross-references: UNIPARC:UPI000017CECA
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of o
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
```

```
Alignment Scores:
Pred. No.:              2.52e-20        Length:         613
Score:                  261.50          Matches:        66
Percent Similarity:     52.5%           Conservative:   8
Best Local Similarity:  46.8%           Mismatches:     26
Query Match:            28.2%           Indels:         41
DB:                     4               Gaps:           5

US-10-607-806-1_COPY_1_500 (1-500) x C40201 (1-613)

Qy          86 GGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACT 145
               ||||||||||||||   |||  ||||||||||||||   :::   |||||||| :
Db         208 GlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyLysGlyGlyTrpIle--M 227

Qy         146 TGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAA 205
               ::|||  ||||||  ||||||||||||||||||||||||||||     |||:::||||
Db         227 etArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPheLeuLeuLysI 247

Qy         206 TACAAAAATTA------GCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGA 259
               |||||||:::      |||  ||||||         |||  |||  :::  ||||
Db         247 leGlnLysIleSerTrpAlaTrpCysGlyArg------Leu***SerHisValArgArgA 265

Qy         260 GGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGC 319
               |||||  |||:::|||   :::|||  |||||||||  ||||||||||||||
Db         265 rgLeuArgGlnGluAsnGlyIleAsnProGlyGlyGlyAlaCysSerGluProArgSerA 285

Qy         320 ACCACTGCACTCCAGCCTGG----------------------------AAAAAAAGGG 349
                 ||||||||||||||||||                            ||||||:::
Db         285 rgHisCysThrProAlaTrpValThrGlu***AspSerValSerLysLysLysLysLysL 305

Qy         350 TAATTAATAACTTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAG------------- 396
                       |||                  |||  |||||||||
Db         305 ys*********Phe------------PhePhePhePhePheGluThrGluSerHisS 321

Qy         397 ----------------------------------------------------CCACCCCCA 405
                                                                   |||||||||
Db         321 erValThrGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProPro 340

RESULT 5
I38022
hypothetical protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C;Accession: I38022
R;Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A;Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A;Reference number: I38021
A;Accession: I38022
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-196 <RES>
A;Cross-references: UNIPARC:UPI000011E1FB; EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PI

Alignment Scores:
Pred. No.:              6.38e-18        Length:         196
Score:                  239.50          Matches:        50
Percent Similarity:     63.2%           Conservative:   10
Best Local Similarity:  52.6%           Mismatches:     20
Query Match:            25.8%           Indels:         15
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_1_500 (1-500) x I38022 (1-196)

Qy         142 CACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATA 201
               ||||||||||:::||||||   |||  ||||||||||||||||||      ||||||:::
Db           3 HisLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIleProSerLeuLeu 22

Qy         202 AAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGG 261
               |||||||||:::||||||      ||||||  ||||||  |||||||||  ||||||
Db          23 LysIleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArg 42

Qy         262 CTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCAC 321
               |||   |||:::|||||||||||||   ||||||   |||||||||||||||:::  |||
Db          43 LeuArgGlnGluAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHis 62

Qy         322 CACTGCACTCCAGCCTGG---------------------------------------- 339
                  |||   |||||||||
Db          63 LeuCysIleProAlaTrpValThrGluGlyAspSerValSerLysGlnAsnLysThrLys 82

Qy         340 ---AAAAAAAGGGTAATTAATAACTTTACTTGCAACCATAGCTGC 381
                  ::::::   :::   ||||||         ::::::||||||
Db          83 AsnGluGlnHisLeuArgAsnAsnThrLysLysSerAsnSerCys 97

RESULT 6
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potential
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Alignment Scores:
Pred. No.:              6.44e-16        Length:         627
Score:                  221.50          Matches:        63
Percent Similarity:     52.5%           Conservative:   11
Best Local Similarity:  44.7%           Mismatches:     30
Query Match:            23.9%           Indels:         37
DB:                     4               Gaps:           7

US-10-607-806-1_COPY_1_500 (1-500) x A40201 (1-627)

Qy         470 CATAACAAGACAGAGGACCAGGCACTGCTCCTAGRTTTAGGCCTGAAGGATGCTAAGTGG 411
               |||:::         |||:::|||      |||   ||||||
Db         291 HisSerSerLeuGlyAspArgAla---ArgLeuArgLeuGly------------------ 303

Qy         410 GTGATTGGGGGTGGCTCAAAGAAGGAGAAGCAGCTATGGTTGCAAGTAAAG--------- 360
                                 ||||||  ||||||         ::::::|||
Db         304 ------------------LysLys***LysGln---------LysIleLysLys****** 314

Qy         359 ---------TTATTAATTACCCTTTTTTTC--------------------------CAG 336
                           :::   |||||||||                           :::
Db         315 ***PhePheTyrPheLeuPheLeuPhePheSerGluThrGluSerArgSerValAlaLys 334

Qy         335 GCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGA 276
               |||||||||||||   :::|||||||||||||   |||||||||   |||
Db         335 AlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProProGlyPheThrPro 354

Qy         275 TTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTGGCT 216
               |||  |||  ||||||      ||||||||||||||:::      ||||||  |||
Db         355 PheSerCysLeuSerLeuProSerSerTrpAspTyrArgArg------ProHisTyrAla 372

Qy         215 AAT------TTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCT 162
               |||||||||||||   ||||||  |||  |||      ||||||||||||
Db         373 ArgLeuIlePheCysIlePheSerArgAspGlyValSerProTrp***ProGlyTrpSer 392

Qy         161 CAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAG 103
               ::::::|||  |||  ||||||  |||      ||||||||||||||:::|||
Db         393 ArgSerProAspLeu--ValIleArgProProTrpProProLysValLeuGlyLeuGln 411
```

```
RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              1.74e-14        Length:         597
Score:                  208.50          Matches:        52
Percent Similarity:     62.1%           Conservative:   7
Best Local Similarity:  54.7%           Mismatches:     24
Query Match:            22.5%           Indels:         13
DB:                     4               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x E40201 (1-597)

Qy        347 CTTTTTTTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCC 288
              |||      |||||||||||||:::|||   ||||||   :::
Db        510 LeuValAlaGlnAlaGlyValArgTrpHisAsnLeuThrAla------------------ 523

Qy        287 AGGTTCAAGTGATTCTTC---------------TGCCCCAGCCTCCTGTGTAGCTGGGAC 243
                  |||      :::                   |||   ||||||      |||||||||
Db        524 AsnPheAlaSerTrpValGlnAlaIleLeuSerCysLeuSerLeuProSerSerTrpAsp 543

Qy        242 TATAGGCAAGTGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTG 183
              ||||||      ||||||      |||||| |||:::||||||:::|||:::   |||
Db        544 TyrArgHisAlaProProArgProAlaAsn-PheIlePheLeuValGluMetGlyPheLe 563

Qy        182 CCATGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAA-CTCGGTCT 124
              ||||||||||||||||||||||||||      |||   |||||||||   |||   ||
Db        563 uHisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAspProProArgLeuGluLe 583

Qy        123 CCCAAAGTGCTGGGAATACAGACATGAGCCACCGCACCTGGCT 81
              |||||||   |||:::|||||||||||||   ||||||||||||
Db        583 uProLysArgTrpAspTyrArgHisGluLeuProHisLeuAla 597

RESULT 8
A40201
artifact-warning sequence (translated ALU class A) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: A40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: A40201
A;Molecule type: DNA
A;Residues: 1-627 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC8
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              2.55e-14        Length:         627
Score:                  207.00          Matches:        48
Percent Similarity:     67.9%           Conservative:   9
Best Local Similarity:  57.1%           Mismatches:     25
Query Match:            22.3%           Indels:         2
DB:                     4               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x A40201 (1-627)

Qy         89 GCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGA 148
              |||   :::|||   |||   |||||||||||||||   :::   |||   ||||||  |
Db          3 AlaArgSerHisAlaCysAsnProSerThrLeuGlyGlyGlnGlyGlyArgIleThr--A 22

Qy        149 GAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATAC 208
              ||:::|||      |||   ||||||   ||||||:::|||   ||||||:::|||||||
Db         22 rgSerGlyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleG 42

Qy        209 AAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGC 268
              |||||::::::|||                |||   |||||||||      ||||||  |
Db         42 lnLysIleSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgG 62

Qy        269 AGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCA 328
              ||:::|||   :::|||   |||||||||   |||||||||||||||      |||||||
Db         62 lnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCysT 82

Qy        329 CTCCAGCC 336
              ||||||||
Db         82 hrProAla 84

RESULT 9
S65657
alpha-1C-adrenergic receptor splice form 2 - human
N;Alternate names: alpha-1C-adrenoceptor isoform 2
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 09-Jul-2004
C;Accession: S65657; S65655
R;Tanaka, T.
submitted to the EMBL Data Library, July 1994
A;Reference number: S65656
A;Accession: S65657
A;Molecule type: mRNA
A;Residues: 1-499 <TAN>
A;Cross-references: UNIPROT:Q13675; UNIPARC:UPI000004E75E; EMBL:D32202; NID:g927208; PID
R;Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.;
FEBS Lett. 363, 256-260, 1995
A;Title: Cloning, functional expression and tissue distribution of human alpha(1C)-adre
A;Reference number: S65654; MUID:95255557; PMID:7737411
A;Accession: S65655
A;Molecule type: mRNA
A;Residues: 424-499 <HIR>
A;Cross-references: UNIPARC:UPI00001778AB; EMBL:D32202
C;Genetics:
A;Gene: GDB:ADRA1C; ADRA1L1
A;Cross-references: GDB:128088; OMIM:104221
A;Map position: 8p21-8p11.2
C;Superfamily: vertebrate rhodopsin
C;Keywords: alternative splicing; neurotransmitter receptor

Alignment Scores:
Pred. No.:              2.76e-12        Length:         499
Score:                  188.50          Matches:        43
Percent Similarity:     65.7%           Conservative:   3
Best Local Similarity:  61.4%           Mismatches:     19
Query Match:            20.3%           Indels:         5
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_1_500 (1-500) x S65657 (1-499)
```

```
Qy        340 TCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCA 281
              :::|||||||||||||||||||   ||||||||||||||||||||   ||||||||||||
Db        430 ThrArgLeuGluCysSerGlyMetIleLeuAlaHisCysAsnLeuArgLeuProGlySer 449

Qy        280 AGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATA----------------GGC 236
              ||||||   |||   ||||||      |||||||||                    |||
Db        450 ArgAspSerProAlaSerAlaSerGlnAlaAlaGlyThrThrGlyAspValProProGly 469

Qy        235 AAGTGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTT 176
              :::   |||         ||||||||||||||||||:::|||      |||   ||||||
Db        470 ArgArgHisGlnAlaGlnLeuIlePheValPheLeuValGluThrGlyPheHisHisVal 489

Qy        175 GGCCAGGCTGGTCTCAAACTCCTGCTCTCA 146
              ||||||      |||   ||||||   |||
Db        490 GlyGlnAspAspLeuAspLeuLeuThrSer 499
```

RESULT 10
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:

Pred. No.:	5.23e-12	Length:	579
Score:	186.00	Matches:	54
Percent Similarity:	54.5%	Conservative:	18
Best Local Similarity:	40.9%	Mismatches:	42
Query Match:	20.0%	Indels:	19
DB:	4	Gaps:	3

US-10-607-806-1_COPY_1_500 (1-500) x D40201 (1-579)

```
Qy         98 CATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGAGCAGGAG 157
              |||   |||   |||||||||||||||         |||      :::     :::||||
Db          6 HisAlaCysAsnProSerThrLeuGlyAlaMetAlaGlyValCysSer--SerSerGlyV 25

Qy        158 TTTGAGACCAGCCTGGCCAACATGGCAAAACACT----ATCTCTAATAAAAATACAAAAA 213
              ||   :::|||||||||||||||   :::|||      :::      |||:::|||||||
Db         25 alGlnGluGlnProGlyGlnHisValGluThrProProLeuGlnLysLysAspThrLysI 45

Qy        214 TTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAG 273
              ||||||||:::|||   |||   |||:::||||||||||||||:::|||||||||
Db         45 leSerGlnIleTrp***HisLysProValValProAlaThrGlnLysAlaGlyAlaGlyG 65

Qy        274 AATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCA 333
              ||||||::: ||||||      ||| :::||||||   |||||||||:::   |||||
Db         65 lySerLeuAsp-LeuGlySerGlnGly-SerValSerHisAspCysThrSerValLeuGl 84

Qy        334 GCCTGGAAAAAAAGGGTAATTAATAACTTTACTTGCAACCATAGCTGCTTCTCCTTCTTT 393
              |||||||:::                  |||:::
Db         84 nProGlyGlnHis----------------ThrLysThrLeuProGlu*********Al 98

Qy        394 GAGCCACCCCCAATCACCCACTTAGCATCCTTCAGGCCTAAAYCTAGGAGCAGTGCCTGG 453
              |||      :::|||                        :::   |||:::
Db         98 aLysHisGlyCysThrProVal--------------------IleProAlaLeu***Gl 111

Qy        454 TCCTCTGTCTTGTTATGACCCCAAGGAA 481
              |||         |||   |||   :::
Db        111 uProTrpGlnValTyrAlaProAlaGln 120
```

RESULT 11
S41044
chromosomal protein - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Oct-2004
C;Accession: S41044
R;Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A;Title: A new chromosomal protein essential for mitotic spindle assembly.
A;Reference number: S41044; MUID:94166884; PMID:8121495
A;Accession: S41044
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-418 <YEO>
A;Cross-references: UNIPROT:P49646; UNIPARC:UPI000013C29A; GB:L26953; NID:g537529; PIDN
C;Superfamily: very very hypothetical protein RMSA-1
C;Keywords: chromosomal protein

Alignment Scores:

Pred. No.:	8.57e-12	Length:	418
Score:	184.00	Matches:	51
Percent Similarity:	48.6%	Conservative:	19
Best Local Similarity:	35.4%	Mismatches:	35
Query Match:	19.8%	Indels:	39
DB:	2	Gaps:	6

US-10-607-806-1_COPY_1_500 (1-500) x S41044 (1-418)

```
Qy         90 CGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAG 149
              |||||||||                  :::||||||   :::   |||   |||||||||
Db        171 ArgTrpLeu-----------------IleTrpGluAlaLysAlaGlyGluSerLeuGlu 184

Qy        150 AGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACT------------------ 191
              ||||||   |||||||||      :::   :::|||:::
Db        185 ValArgSerSerArgProAlaSerGlnSerArgArgAsnSerValSerThrLysAsnIle 204

Qy        192 ------------ATCTCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGG--------- 230
                         :::|||   ||||||   ||||||||||||   |||:::
Db        205 LysIleSerProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGly 224

Qy        231 ------CACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAGAAGAATCACTTGAA 284
                   ||||||   :::|||||||   |||:::::|||   |||   |||      :::
Db        225 GlyValHisLeu---LeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAsp 243

Qy        285 CCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACTCCAGCCTGG----- 339
              ||| ||||||   ||||||:::|||   |||      ||||||||||||||||||
Db        244 Pro-GlyGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValTh 263

Qy        340 ---------------------------------AAAAAAAGGGTAATTAATAACTTTAC 365
                                                ||||||:::              ||
Db        263 rSerGluThrValSerLysLysLysLysLysLysLysLysLysLysThrArgValLeuTh 283

Qy        366 TTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCACTTAGCATCCTT 425
              ||||   ::::::         |||               :::|||   :::      ::
Db        283 rCysIleAsnAlaSerThrLeuPheHis---------ValLeuThrArgIlePheCysTy 300

Qy        426 CAGGCCTAAA 435
              ::::   |||
Db        300 rLysGlnLys 303
```

RESULT 12
E40201
artifact-warning sequence (translated ALU class E) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: E40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201

```
A;Accession: E40201
A;Molecule type: DNA
A;Residues: 1-597 <CLA>
A;Cross-references: UNIPARC:UPI000017CECC
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              7.52e-11      Length:         597
Score:                  175.50        Matches:        59
Percent Similarity:     50.4%         Conservative:   7
Best Local Similarity:  45.0%         Mismatches:     50
Query Match:            18.9%         Indels:         16
DB:                     4             Gaps:           3

US-10-607-806-1_COPY_1_500 (1-500) x E40201 (1-597)

Qy          82 GCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAG-ACCGAGTTGGGTGGA 140
               ||||||||||||  |||||||||  ||||||  ||||||   :::      ||||||
Db         101 AlaArgCysGlySerSerCysLeu***SerGlnArgPheGlySerSerArgArgGlyGly 120

Qy         141 TCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAAT 200
               |||   |||       ||||||||||||||||   |||||||||:::|||   ||||||
Db         121 SerProGluValGlySerLeuArgProAla***ProThrTrpArgAsnProIleSerThr 140

Qy         201 AAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAG 260
               ||||||    |||||||||      ||||||   ||||||   |||||||||      ||
Db         141 LysAsnIle-LysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGlnLeuLeuGlyAr 160

Qy         261 GCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCA 320
               ||||   |||:::      :::                                ||||||
Db         160 gLeuArgGlnGlu---ArgIleAla***ThrGlnGluAlaLysPheAlaValArgLeuCy 179

Qy         321 CCACTGCACTCCAGCCTGG----------------------------AAAAAAAGGGTAAT 353
                |||   ||||||||||||                           ||||||:::
Db         179 sHisArgThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLysLysLys***** 199

Qy         354 TAATAACTTTACTTGCAACCATAGCTGCTTCTCCTTCTTTGAGCCACCCCCAATCACCCA 413
                        |||:::||| :::||||||             |||
Db         199 ****ProGlyValValAlaHisAlaCys-AsnProSerAlaLeuGlyValGlnGlyGlyA 219

Qy         414 CTTAGCATCCTTCAGGCCTAAAYCTAGGA 442
                  |||         |||   ||||||
Db         219 laAspHis---------LeuArgLeuGly 225

RESULT 13
D40201
artifact-warning sequence (translated ALU class D) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: D40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: D40201
A;Molecule type: DNA
A;Residues: 1-579 <CLA>
A;Cross-references: UNIPARC:UPI000017CECB
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              2.67e-10      Length:         579
Score:                  170.50        Matches:        47
Percent Similarity:     63.9%         Conservative:   6
Best Local Similarity:  56.6%         Mismatches:     25
Query Match:            18.4%         Indels:         6
DB:                     4             Gaps:           1

US-10-607-806-1_COPY_1_500 (1-500) x D40201 (1-579)

Qy         339 CCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAA 280
               ||||||||||||  ||||||||||||||||||| :::|||   ||||||  ||||
Db         301 ProGlyTrpSerThrGluValGlnSerTrpLeuThr-GluPro-***LeuProArgSerS 320

Qy         279 GTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCT 220
               |||||     |||||||||      ||||||||||||   |||   |||:::      |
Db         320 erAspProProAlaProAlaPhe***ValAlaGlyThrThrGlyLeuCysTyrGlnIleT 340

Qy         219 GGCTAATTTTTGTATT----TTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGT 164
               |||||||||||||||      |||   |||      |||      |||   ||||||   |
Db         340 rpLeuIlePheValSerPhePheCysArgGlyGlyValSerThrCysCysProGlyCysS 360

Qy         163 CTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACA 104
               ||   ||||||   |||:::   :::  |||      :::|||||||||||||||:::||
Db         360 er***ThrProGluLeuGluHisThr--ProAlaMetAlaProLysValLeuGlyLeuGl 379

Qy         103 G 103
               |
Db         379 n 379

RESULT 14
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked re
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: UNIPROT:Q07826; UNIPARC:UPI000006E23F; GB:S58722; NID:g299470; PIDN
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.:              7.77e-10      Length:         100
Score:                  166.00        Matches:        35
Percent Similarity:     79.6%         Conservative:   4
Best Local Similarity:  71.4%         Mismatches:     9
Query Match:            17.9%         Indels:         1
DB:                     2             Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x A46010 (1-100)

Qy         338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
               :::|||||||||||||||||:::|||||||||||:::   ||||||  ||||||||||
Db          15 GluAlaGlyValGlnTrpCysAspLeuGlySerLeuLysSerProPro-ProGlySerSe 34

Qy         278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
               |||||||   |||   ||||||   |||||||||      |||      |||||||||
Db          34 rAspSerProAlaSerAlaSerArgValAlaGlyIleThrGlyMetHisHisHisThrGl 54

Qy         218 GCTAATTTTTGTATTTTTATTAGAG 194
                 ||||||||||||||||||:::|||
Db          54 nLeuIlePheValPheLeuValGlu 62

RESULT 15
B40201
```

```
artifact-warning sequence (translated ALU class B) - human
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C;Accession: B40201
R;Claverie, J.M.
personal communication, 1992
A;Reference number: A40201
A;Accession: B40201
A;Molecule type: DNA
A;Residues: 1-301 <CLA>
A;Cross-references: UNIPARC:UPI000017CEC9
R;Claverie, J.M.
Genomics 12, 838-841, 1992
A;Title: Identifying coding exons by similarity search: Alu-derived and other potentiall
A;Reference number: A40200; MUID:92241891; PMID:1572661
A;Contents: annotation
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
C;Comment: Any significant similarity of a predicted protein sequence to a portion of th
.

Alignment Scores:
Pred. No.:              6.19e-09        Length:         301
Score:                  158.00          Matches:        42
Percent Similarity:     58.8%           Conservative:   15
Best Local Similarity:  43.3%           Mismatches:     31
Query Match:            17.0%           Indels:         9
DB:                     4               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x B40201 (1-301)

Qy        362 AAGTTATTAATTACCCTTTTTTTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTG 303
              :::|||::::::      |||::::::|||||||||      |||
Db        205 GlnLeuIleLeuPheTyrPheTyrArgAlaGlyValLeuLeuCysCysProSerTrpSer 224

Qy        302 CAACCTCCACCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGAC 243
                  |||      :::         :::|||     |||:::          ||||||
Db        225 ***ThrProGlyLeuLysGlnSerSerTyrPhe------SerIleProLysCysTrpAsp 242

Qy        242 TATAGGCAAGTGCCACCACACCTGGCTAAT-----------------TTTTGTATTTTT 201
              ||||||      ||||||      |||                  :::   ||||||
Db        243 TyrArgCys***ProProCysProAla*********ProSer***PheTyrPheIlePhe 262

Qy        200 ATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGA 141
              ||| |||:::      |||:::|||   ||||||||||||:::||||||:::||||||::
Db        263 Ile-GluLeuGlySerCysTyrValAlaGlnAlaGlyLeuGluLeuLeuValSerSerAs 282

Qy        140 TCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACATGAGCCAC 92
              :||||||   |||   ||||||||||||||   ||||||:::||||||
Db        282 nProProThrSerAlaSerGlnSerAlaGlyIleThrAspValSerHis 298


Search completed: February  8, 2006, 10:03:29
Job time : 23.4044 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 8, 2006, 09:18:47 ; Search time 24.2672 Seconds
(without alignments)
2907.338 Million cell updates/sec

Title: US-10-607-806-1_COPY_1_500
Perfect score: 928
Sequence: 1 gacctacctcgacctttgtg.....acccacataagagggactga 500

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10607806/runat_08022006_091819_26843/app_query.fasta_1
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p
-USER=US10607806_@CGN_1_1_694_@runat_08022006_091819_26843 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	359	38.7	593	1	ALU7_HUMAN	P39194 homo sapien
c	2	343	37.0	593	1	ALU6_HUMAN	P39193 homo sapien
	3	322	34.7	593	1	ALU7_HUMAN	P39194 homo sapien
c	4	318	34.3	260	2	Q6ZQR8_HUMAN	Q6zqr8 homo sapien
	5	317	34.2	136	2	Q96NR6_HUMAN	Q96nr6 homo sapien
c	6	314	33.8	158	2	Q8NAL9_HUMAN	Q8nal9 homo sapien
c	7	311.5	33.6	138	2	Q8N874_HUMAN	Q8n874 homo sapien
c	8	311	33.5	140	2	Q6ZP99_HUMAN	Q6zp99 homo sapien
c	9	310	33.4	124	2	Q6ZVF6_HUMAN	Q6zvf6 homo sapien
c	10	307.5	33.1	202	2	Q6ZUA4_HUMAN	Q6zua4 homo sapien
c	11	307	33.1	585	1	ALU5_HUMAN	P39192 homo sapien
c	12	305.5	32.9	122	2	Q6ZVH4_HUMAN	Q6zvh4 homo sapien
	13	304	32.8	593	1	ALU6_HUMAN	P39193 homo sapien
c	14	301.5	32.5	166	2	Q6ZUC0_HUMAN	Q6zuc0 homo sapien
c	15	301	32.4	591	1	ALU1_HUMAN	P39188 homo sapien
c	16	300	32.3	94	2	Q8NED3_HUMAN	Q8ned3 homo sapien
c	17	297	32.0	171	2	Q9H728_HUMAN	Q9h728 homo sapien
	18	297	32.0	591	1	ALU1_HUMAN	P39188 homo sapien
	19	293.5	31.6	591	1	ALU8_HUMAN	P39195 homo sapien
c	20	293	31.6	160	2	Q6ZWD2_HUMAN	Q6zwd2 homo sapien
	21	293	31.6	218	2	Q8NAC7_HUMAN	Q8nac7 homo sapien
c	22	290	31.2	165	2	Q6ZS53_HUMAN	Q6zs53 homo sapien
c	23	290	31.2	165	2	Q6ZT71_HUMAN	Q6zt71 homo sapien
c	24	289	31.1	115	2	Q9N083_MACFA	Q9n083 macaca fasc
c	25	288.5	31.1	137	2	Q6ZS97_HUMAN	Q6zs97 homo sapien
c	26	286	30.8	127	2	Q6ZVW2_HUMAN	Q6zvw2 homo sapien
c	27	285	30.7	130	2	Q6ZSV2_HUMAN	Q6zsv2 homo sapien
c	28	284.5	30.7	147	2	Q6ZUA2_HUMAN	Q6zua2 homo sapien
c	29	284.5	30.7	591	1	ALU8_HUMAN	P39195 homo sapien
c	30	284	30.6	220	2	Q5SVL1_HUMAN	Q5svl1 homo sapien
c	31	284	30.6	587	1	ALU2_HUMAN	P39189 homo sapien
c	32	283	30.5	138	2	Q6ZUF3_HUMAN	Q6zuf3 homo sapien
	33	282	30.4	124	2	Q6ZU63_HUMAN	Q6zu63 homo sapien
	34	282	30.4	585	1	ALU5_HUMAN	P39192 homo sapien
c	35	281	30.3	127	2	Q9H743_HUMAN	Q9h743 homo sapien
c	36	281	30.3	138	2	Q6ZUU5_HUMAN	Q6zuu5 homo sapien
c	37	281	30.3	150	2	Q6ZPA0_HUMAN	Q6zpa0 homo sapien
c	38	279.5	30.1	125	2	Q6ZTE1_HUMAN	Q6zte1 homo sapien
c	39	279	30.1	156	2	Q8NBH4_HUMAN	Q8nbh4 homo sapien
c	40	279	30.1	164	2	Q6ZUK0_HUMAN	Q6zuk0 homo sapien
c	41	278.5	30.0	137	2	Q6ZP21_HUMAN	Q6zp21 homo sapien
c	42	278	30.0	118	2	Q9H387_HUMAN	Q9h387 homo sapien
	43	278	30.0	143	2	Q6ZVQ4_HUMAN	Q6zvq4 homo sapien
c	44	277	29.8	124	2	Q6ZSN0_HUMAN	Q6zsn0 homo sapien
c	45	277	29.8	132	2	Q6ZNG3_HUMAN	Q6zng3 homo sapien

ALIGNMENTS

RESULT 1
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely
RT connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall

```
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:                5.18e-30        Length:          593
Score:                    359.00          Matches:         72
Percent Similarity:       66.1%           Conservative:    12
Best Local Similarity:    56.7%           Mismatches:      37
Query Match:              38.7%           Indels:          6
DB:                       1               Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x ALU7_HUMAN (1-593)

Qy        461 ACAGAGGACCAGGCACTGCTCCTAGRTTTAGGCCTGAAGGATGCTAAGTGGGTGATTGGG 402
              :::   |||:::|||     |||     |||   ::::::        |||
Db        275 SerArgAspArgAlaThrAlaLeuGlnProGlyGlnGlnGluArgAsnSerVal------ 292

Qy        401 GGTGGCTCAAAGAAGGAGAAGCAGCTATGGTTGCAAGTAAAGTTATTAATTACCCTTTTT 342
                  |||||||||          :::|||      :::          |||
Db        293 ------SerLysLys*********PhePheLeu------ArgArgSerPheAlaLeuVal 308

Qy        341 TTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTC 282
                |||||||||||||||||   :::|||||||||||||||||||||||||||    |||
Db        309 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProProProProGlyPhe 328

Qy        281 AAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACAC 222
              |||   |||   |||   ||||||      |||||||||||||:::   ||||||
Db        329 LysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProProProArg 348

Qy        221 CTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCT 162
```

```
                |||||||||||||||||   ||||||   |||   ||||||||||||||||||||
Db        349 ProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSer 368

Qy        161 CAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGA 102
              :::||||||   |||:::   ||||||:::|||||||||||||||||||:::||||||
Db        369 ArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspTyrArg 388

Qy        101 CATGAGCCACCGCACCTGGCT 81
                  |||||||||      |||
Db        389 ArgGluProProArgProAla 395

RESULT 2
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
```

```
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:                2.94e-28        Length:          593
Score:                    343.00          Matches:         69
Percent Similarity:       63.8%           Conservative:    12
Best Local Similarity:    54.3%           Mismatches:      40
Query Match:              37.0%           Indels:          6
DB:                       1               Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x ALU6_HUMAN (1-593)

Qy        461 ACAGAGGACCAGGCACTGCTCCTAGRTTTAGGCCTGAAGGATGCTAAGTGGGTGATTGGG 402
              :::   |||:::|||:::   |||      |||   ::::::          |||
Db        275 SerArgAspArgAlaIleAlaLeuGlnProGlyGlnGlnGluArgAsnSerVal------ 292

Qy        401 GGTGGCTCAAAGAAGGAGAAGCAGCTATGGTTGCAAGTAAAGTTATTAATTACCCTTTTT 342
                  |||||||||           :::|||      :::          |||
Db        293 ------SerLysLys*********PhePheLeu------ArgArgSerPheAlaLeuVal 308

Qy        341 TTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTC 282
              |||||||||||||||||   :::|||||||||   |||||||||||||   |||
Db        309 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerProGlnProProProProGlyPhe 328

Qy        281 AAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACAC 222
              |||   |||   |||   ||||||      ||||||||||||||      ||||||
Db        329 LysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArg 348

Qy        221 CTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCT 162
              |||||||||||||||||   ||||||   |||   |||||||||   |||||||||
Db        349 ProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSer 368

Qy        161 CAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGA 102
              :::||||||   |||:::   |||   :::|||||||||||||||||||:::||||||
Db        369 ArgThrProAspLeuArg***SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArg 388

Qy        101 CATGAGCCACCGCACCTGGCT 81
                |||||||||      |||
Db        389 ArgGluProProArgProAla 395

RESULT 3
ALU7_HUMAN
ID   ALU7_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39194;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SQ sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14573; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
```

```
SQ   SEQUENCE   593 AA;  64455 MW;  54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.:                5.9e-26        Length:          593
Score:                    322.00         Matches:         60
Percent Similarity:       76.1%          Conservative:    7
Best Local Similarity:    68.2%          Mismatches:      21
Query Match:              34.7%          Indels:          0
DB:                       1              Gaps:            0

US-10-607-806-1_COPY_1_500 (1-500) x ALU7_HUMAN (1-593)

Qy          84 CAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCA 143
               :::   |||||||||      |||   ||||||||||||||||   |||   |||||||||
Db           2 ArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySer 21

Qy         144 CTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAA 203
                 |||   ||||||   ||||||||||||||||||||   |||   :::|||   |||
Db          22 ProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThrLys 41

Qy         204 AATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCT 263
               ||||||||||||||:::   ||||||      |||::::::||||||||:::||||||
Db          42 AsnThrLysIleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAla 61

Qy         264 GGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCA 323
                 |||   ||||||||||||||||||||   |||||||||   |||:::|||||||||
Db          62 GluAlaGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaPro 81

Qy         324 CTGCACTCCAGCCTGGAAAAAAAG 347
               ||||||||||||||||      |||
Db          82 LeuHisSerSerLeuGlyAsnLys 89

RESULT 4
Q6ZQR8_HUMAN
ID   Q6ZQR8_HUMAN PRELIMINARY;      PRT;   260 AA.
AC   Q6ZQR8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ45973.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Placenta;
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA   Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA   Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK128808; BAC87615.1; -; mRNA.
SQ   SEQUENCE   260 AA;  28819 MW;  5D5C34F441BC5994 CRC64;

Alignment Scores:
Pred. No.:                1.34e-25       Length:          260
Score:                    318.00         Matches:         61
Percent Similarity:       63.1%          Conservative:    4
Best Local Similarity:    59.2%          Mismatches:      22
Query Match:              34.3%          Indels:          16
DB:                       2              Gaps:            2

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZQR8_HUMAN (1-260)

Qy         338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
               |||||||||   ||||||||||||||   ||||||||||||||||||||      |||
Db          36 GlnAlaGlyAlaGlnTrpCysAsnLeuSerSerLeuGlnProProProProGluLeuLys 55

Qy         278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
               |||||||||   |||||||||   ||||||||||||||||      ||||||
Db          56 GlyPhePheCysLeuSerLeuLeuThrSerTrpAspTyrArgHisAlaProProArgPro 75

Qy         218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
               ||||||||||||||||||   ||||||   |||                ||||||||||:::
Db          76 AlaAsnPheCysIlePheSerArgAspGlyValSer---------ProGlyTrpSerArg 92

Qy         158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
               |||      |||:::   ||||||   |||   ||||||||||||||:::||||||
Db          93 ThrSerAspLeuArgTrpSerThrHisLeuCysLeuProLysCysTrpAspTyrArgArg 112

Qy          98 GAGCCACCGCACCTGGCTAATGAGTTACTATTAACTGAGCGCTAACTCCATCCCAGGTCC 39
               ||||||                                          |||||||||   :::
Db         113 GluPro----------------------------------------LeuHisProAlaThr 119

Qy          38 CATATGCTA 30
               |||   |||
Db         120 HisAlaLeu 122

RESULT 5
Q96NR6_HUMAN
ID   Q96NR6_HUMAN PRELIMINARY;      PRT;   136 AA.
AC   Q96NR6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ30278.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Cerebellum;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK054840; BAB70813.1; -; mRNA.
SQ   SEQUENCE   136 AA;  14980 MW;  D2336B649A110163 CRC64;

Alignment Scores:
```

Pred. No.:	1.48e-25	Length:	136
Score:	317.00	Matches:	61
Percent Similarity:	72.8%	Conservative:	6
Best Local Similarity:	66.3%	Mismatches:	24
Query Match:	34.2%	Indels:	1
DB:	2	Gaps:	0

US-10-607-806-1_COPY_1_500 (1-500) x Q96NR6_HUMAN (1-136)

```
Qy         75 CTCATTAGCCAGGTGCGGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTG 134
              :::|||         :::||||||      |||   ||||||||||||||||||||
Db         28 ValIleGlyLeuAlaGlnTrpLeuThrProValIleProAlaLeuTrpGluThrGluAla 47

Qy        135 GGTGGATCACTTGAGAGCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATC 194
              ||||||||||||||   ||||||   |||||||||||||||||||:::|||   |||
Db         48 GlyGlySerLeuGluValArgSerSerArgProAlaTrpProThrTrpArgAsnProIle 67

Qy        195 TCTAATAAAAATACAAAAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACA 254
              |||   ||||||||||||||||||||   ||||||   :::|||:::|||   ||||||
Db         68 SerThrLysAsnThrLysIleSerGlnThrTrpTrpArgMetProValValLeuAlaThr 87

Qy        255 CAGGAGGCTGGGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAG 314
              ||||||   |||   ||||||:::||||||||||||   ||||||||||   ||
Db         88 TrpGluAlaGluAlaGlyGluSer-ValAsnLeuGlyGlyArgGlyCysSerGluLeuAr 107

Qy        315 ATTGCACCACTGCACTCCAGCCTGGAAAAAAAGG 348
              |         ||||||||||||||||      |||
Db        107 gArgCysArgCysThrProAlaTrpAlaThrArg 118
```

RESULT 6
Q8NAL9_HUMAN
ID Q8NAL9_HUMAN PRELIMINARY; PRT; 158 AA.
AC Q8NAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK092450; BAC03893.1; -; mRNA.
DR Ensembl; ENSG00000198057; Homo sapiens.
SQ SEQUENCE 158 AA; 17414 MW; E151503015F2FE34 CRC64;

Pred. No.:	3.27e-25	Length:	158
Score:	314.00	Matches:	67
Percent Similarity:	49.3%	Conservative:	5
Best Local Similarity:	45.9%	Mismatches:	13
Query Match:	33.8%	Indels:	61
DB:	2	Gaps:	1

US-10-607-806-1_COPY_1_500 (1-500) x Q8NAL9_HUMAN (1-158)

```
Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||||   ||||||||||||||||||   ||||||   ||||||
Db          8 GlnAlaGlyValGlnTrpCysGlyLeuGlySerLeuGlnProLeuProProGlyPheLys 27

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
              |||||||||   |||||||||   |||||||||||||:::|||||||||   |||
Db         28 ArgPhePheCysLeuSerLeuLeuSerSerTrpAspTyrArgArgValProProProLeu 47

Qy        218 GCTAATTTTTGTATTTTT---------------------------------------------- 201
              ||||||||||||||||||
Db         48 AlaAsnPheCysIlePhePheSerPhePhePheGluLysGluSerLeuSerValThrGln 67

Qy        201 ---------------------------------------------------------------- 201

Db         68 AlaGlyValGlnTrpHisAspLeuGlySerLeuGlnAlaAlaProProGlyPheThrPro 87

Qy        201 ---------------------------------------------------------------- 201

Db         88 PheSerCysLeuSerLeuProSerSerTrpAsnTyrArgArgProProProCysProAla 107

Qy        200 ------------------ATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
                                 ::: |||:::   |||||||||:::|||||||||||||||::
Db        108 AsnPhePheValPheLeuVal-GluMetGlyPheCysHisIleGlyGlnAlaGlyLeuGl 127

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              :||||||   ||||||||||||||   |||   ||||||||||||||   |||   ||
Db        127 uLeuLeuThrSerSerAspProProThrSerAlaSerGlnSerAlaGlyIleThrGlyMe 147

Qy         98 GAGCCACCGCACCTGG 83
              ||||||||||||||||
Db        147 tSerHisArgThrTrp 152
```

RESULT 7
Q8N874_HUMAN
ID Q8N874_HUMAN PRELIMINARY; PRT; 138 AA.
AC Q8N874;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ39895.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Spleen;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,

```
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
DR   EMBL; AK097214; BAC04976.1; -; mRNA.
DR   Ensembl; ENSG00000198424; Homo sapiens.
SQ   SEQUENCE   138 AA;  15427 MW;  685645DF92E31994 CRC64;

Alignment Scores:
Pred. No.:              5.96e-25        Length:         138
Score:                  311.50          Matches:        65
Percent Similarity:     70.7%           Conservative:   5
Best Local Similarity:  65.7%           Mismatches:     27
Query Match:            33.6%           Indels:         2
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_1_500 (1-500) x Q8N874_HUMAN (1-138)

Qy        377 CTATGGTTGCAAGTAAAGTTATTAATTACCCTTTTTTTCCAGGCTGGAGTGCAGTGGTGC 318
              |||:::|||   :::           ||||||      |||   ||||||||||||
Db         12 LeuPheLeuPheLeuSerTrpSerPheThrLeuValAlaGlnThrGlyValGlnTrpArg 31

Qy        317 AATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTC 258
              ||||||||||||||||||||||||||||||||   ||||||   |||   |||      |||
Db         32 AsnLeuGlySerLeuGlnProProProProGlyPheLysArgPheSerCysLeuGlyLeu 51

Qy        257 CTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTGGCTAATTTTTGTATTTTTATT 198
              |||||||||||||||       ||||||   ||||||||||||||   |||
Db         52 ProSerSerTrpAspTyrArg-His---HisHisAlaTrpLeuIlePheValLeuLeuAl 70

Qy        197 AGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCC 138
               |||:::   ||||||||||||||||||||||||||:::||||||   ||||||:::||
Db         70 aGluMetGlyPheCysHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerSerGluPr 90

Qy        137 ACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACATGAGCCACCGCACCTGG 83
              ||||   |||   ||||||||||||||   |||   |||||||||   ||||||
Db         90 oProAlaSerAlaSerGlnSerAlaGlyIleThrGlyMetSerHisHisThrTrp 108

RESULT 8
Q6ZP99_HUMAN
ID   Q6ZP99_HUMAN PRELIMINARY;      PRT;   140 AA.
AC   Q6ZP99;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ26180.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Adrenal gland;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Suzuki Y., Hata H.,
RA   Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA   Nagai K., Isogai T., Sugano S.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK129691; BAC85220.1; -; mRNA.
SQ   SEQUENCE   140 AA;  15350 MW;  27A3F9D2C23FF9DD CRC64;

Alignment Scores:
Pred. No.:              6.79e-25        Length:         140
Score:                  311.00          Matches:        60
Percent Similarity:     72.2%           Conservative:   5
Best Local Similarity:  66.7%           Mismatches:     25
Query Match:            33.5%           Indels:         0
DB:                     2               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZP99_HUMAN (1-140)

Qy        352 TTACCCTTTTTTTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCAC 293
              |||            :::|||||||||   |||   |||   ||||||||||||||||||
Db         30 LeuAlaLeuLeuProLysLeuGluCysArgGlyThrIleSerAlaHisCysAsnLeuHis 49

Qy        292 CTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAG 233
              ||||||||||||||||       |||   ||||||   ||||||||||   |||
Db         50 LeuProGlySerSerAspPheProAlaSerAlaSerGlnValAlaGlyThrThrGlyAla 69

Qy        232 TGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGC 173
              |||||||||   ||||||||||||||||||||:::|||      |||   |||||||||
Db         70 CysHisHisAlaTrpLeuIlePheValPheLeuValGluAlaGlyPheHisHisValGly 89

Qy        172 CAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCT 113
              |||   ||||||:::|||||   |||:::|||||||||          ||||||||||||
Db         90 GlnAspGlyLeuGluLeuLeuThrSerAsnAspProProThrLeuAlaSerGlnSerAla 109

Qy        112 GGGAATACAGACATGAGCCACCGCACCTGG 83
              |||   |||   :::|||||||||   |||
Db        110 GlyIleThrGlyValSerHisArgAlaTrp 119

RESULT 9
Q6ZVF6_HUMAN
ID   Q6ZVF6_HUMAN PRELIMINARY;      PRT;   124 AA.
AC   Q6ZVF6;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ42633.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Cerebellum;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK124624; BAC85905.1; -; mRNA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   124 AA;  14173 MW;  108AA891087A31B4 CRC64;
```

```
Alignment Scores:
Pred. No.:                8.49e-25        Length:         124
Score:                    310.00          Matches:        59
Percent Similarity:       80.0%           Conservative:   5
Best Local Similarity:    73.8%           Mismatches:     14
Query Match:              33.4%           Indels:         2
DB:                       2               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZVF6_HUMAN (1-124)

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||||||::::::   ||||||||||||   ||||||   ||||||
Db         46 GlnAlaGlyValGlnTrpCysAspIleSerSerLeuGlnProLeuProProGlyPheLys 65

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
              |||||||||   |||||||||   ||||||||||||||||||:::||||||   |||
Db         66 GlnPhePheCysLeuSerLeuLeuSerSerTrpAspTyrArgGlnLeuProProCysLeu 85

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||||||   ||||||||||   |||||||||   ||||||||||||
Db         86 AlaAsnPheCysIlePheSerArgAspSerValSerProCysTrpSerGlyTrpSerGln 105

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAG 103
              ||||||   |||  :::|||   |||   |||   ||||||||||||||:::|||
Db        106 ThrProAspLeu--LeuIleArgProProArgProProLysValLeuGlyLeuGln 123

RESULT 10
Q6ZUA4_HUMAN
ID   Q6ZUA4_HUMAN PRELIMINARY;      PRT;   202 AA.
AC   Q6ZUA4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ43866.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Testis;
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA   Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA   Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK125854; BAC86322.1; -; mRNA.
SQ   SEQUENCE   202 AA;  22509 MW;  7C402B7E86DD01FA CRC64;

Alignment Scores:
Pred. No.:                1.79e-24        Length:         202
Score:                    307.50          Matches:        68
Percent Similarity:       66.4%           Conservative:   5
Best Local Similarity:    61.8%           Mismatches:     26
Query Match:              33.1%           Indels:         11
DB:                       2               Gaps:           2

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZUA4_HUMAN (1-202)

Qy        391 AGAAGGAGAAGCAGCTATGGTTGCAAGTAAAGTTATTAATTACCCTTTTTTTCC------ 338
              |||   :::   ||||||                  :::   |||   |||||||||
Db         30 ArgSerGlnAspSerTyrAsn---------ProPheIleLeuLeuPhePheSerArgArg 46

Qy        337 ------------------AGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTC 296
                               ||||||||||||||||||   ||||||||||||||||||||
Db         47 SerLeuSerLeuSerProArgLeuGluCysSerGlyValIleLeuAlaHisCysAsnLeu 66

Qy        295 CACCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGC 236
              |||||||||||||||||||||||   |||   ||||||   |||||||||
Db         67 HisLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaGlyIleThrVal 86

Qy        235 AAGTGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTT 176
              ||||||   ||||||||||||||||||||||:::|||   |||   ||||||
Db         87 AlaArgHisHisAlaTrpLeuIlePheValPheLeuValGluThrGlyPheHisHisVal 106

Qy        175 GGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGT 116
              ||||||||||||||:::||||||   |||   |||||||||   |||   |||||||||
Db        107 GlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspProProThrSerAlaSerGlnSer 126

Qy        115 GCTGGGAATACAGACATGAGCCACCGCACC 86
              ||||||   |||   :::||||||   |||
Db        127 AlaGlyIleThrGlyValSerHisHisThr 136

RESULT 11
ALU5_HUMAN
ID   ALU5_HUMAN     STANDARD;      PRT;   585 AA.
AC   P39192;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SC sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
```

```
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14571; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     95       Frame-1.
FT   REGION       99    193       Frame-2.
FT   REGION      197    291       Frame-3.
FT   REGION      295    389       Frame-4.
FT   REGION      393    487       Frame-5.
FT   REGION      491    585       Frame-6.
SQ   SEQUENCE   585 AA;  63994 MW;  46EE8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.:              2.6e-24        Length:        585
Score:                  307.00         Matches:       61
Percent Similarity:     72.8%          Conservative:  6
Best Local Similarity:  66.3%          Mismatches:    23
Query Match:            33.1%          Indels:        2
DB:                     1              Gaps:          0

US-10-607-806-1_COPY_1_500 (1-500) x ALU5_HUMAN (1-585)

Qy        353 ATTACCCTTTTTTTCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCA 294
              ::::::|||      |||||||||||||||||||   :::|||||||||||||||||||||
Db        398 ValSerLeuCysArgGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProPro 417

Qy        293 CCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAA 234
              ||||||  ||||||   |||  |||  ||||||      |||||||||||||||:::
Db        418 ProProGlyPheLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArg 437

Qy        233 GTGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGG 174
                ||||||      |||||||||||||||||||   ||||||   |||   |||||||||
Db        438 AlaProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerProCysTrp 457

Qy        173 CCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGC 114
              ||||||||||||:::  |||        ||||||   |||   |||   |||||||||
Db        458 ProGlyTrpSerArg--SerLeuAspLeuValIleArgProProArgProProLysValL 477

Qy        113 TGGGAATACAGACATGAGCCACCGCACCTGGC 82
              |||||:::|||      ||||||||||||||
Db        477 euGlyLeuGlnAla***AlaThrAlaProGly 487

RESULT 12
Q6ZVH4_HUMAN
ID   Q6ZVH4_HUMAN PRELIMINARY;      PRT;   122 AA.
AC   Q6ZVH4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ42572.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Cerebellum;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK124563; BAC85887.1; -; mRNA.
SQ   SEQUENCE   122 AA;  13547 MW;  39472EF24DED96D3 CRC64;

Alignment Scores:
Pred. No.:              2.64e-24       Length:        122
Score:                  305.50         Matches:       63
Percent Similarity:     65.3%          Conservative:  1
Best Local Similarity:  64.3%          Mismatches:    24
Query Match:            32.9%          Indels:        10
DB:                     2              Gaps:          1

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZVH4_HUMAN (1-122)

Qy        347 CTTTTTTTC-------------------------CAGGCTGGAGTGCAGTGGTGCAAT 315
              |||||||||                         ||||||||||||||||||||:::
Db         25 LeuPhePheIleGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpCysAsp 44

Qy        314 CTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTCTTCTGCCCCAGCCTCCTG 255
              |||  ||||||||||||||||||||  ||||||  |||  |||  ||||||
Db         45 LeuSerSerLeuGlnProProProProGlyPheLysArgPheSerCysLeuSerLeuPro 64

Qy        254 TGTAGCTGGGACTATAGGCAAGTGCCACCACACCTGGCTAATTTTTGTATTTTTATTAGA 195
                 ||||||||||||||      ||||||  |||      ||||||||||||  |||
Db         65 SerSerTrpAspTyrArgCysProProProArgLeuValThrPheCysIlePheSerArg 84

Qy        194 GATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGCTCTCAAGTGATCCACC 135
              |||  |||      ||||||||||| ||||||||||||||  |||  |||  ||
Db         85 AspArgValSerSerCysTrpProGly-GlyLeuLysLeuLeuThrSerGlyAspLeuPr 104

Qy        134 CAACTCGGTCTCCCAAAGTGCTGGGAATACAGACATGAGCCACCGCACCTGG 83
              |  |||  ||||||||||||||  |||  ||||||||  ||||||
Db        104 oAlaSerAlaSerGlnSerAlaGlyValThrGlyMetSerHisHisThrTrp 121

RESULT 13
ALU6_HUMAN
ID   ALU6_HUMAN     STANDARD;      PRT;   593 AA.
AC   P39193;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily SP sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
```

```
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14572; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR   InterPro; IPR002218; GIDA.
DR   ProDom; PD003738; GIDA; 1.
KW   Hypothetical protein.
FT   REGION        1     97       Frame-1.
FT   REGION      101    196       Frame-2.
FT   REGION      200    295       Frame-3.
FT   REGION      299    395       Frame-4.
FT   REGION      399    494       Frame-5.
FT   REGION      498    593       Frame-6.
SQ   SEQUENCE   593 AA;  64640 MW;  136EF344AACD12A2 CRC64;

Alignment Scores:
Pred. No.:              5.55e-24        Length:         593
Score:                  304.00          Matches:        60
Percent Similarity:     74.4%           Conservative:   4
Best Local Similarity:  69.8%           Mismatches:     22
Query Match:            32.8%           Indels:         0
DB:                     1               Gaps:           0

US-10-607-806-1_COPY_1_500 (1-500) x ALU6_HUMAN (1-593)

Qy          91 GGTGGCTCATGTCTGTATTCCCAGCACTTTGGGAGACCGAGTTGGGTGGATCACTTGAGA 150
               |||||||||   |||   ||||||||||||||||||||          ||||||||||||
Db         104 GlyGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAspHisLeuArg 123

Qy         151 GCAGGAGTTTGAGACCAGCCTGGCCAACATGGCAAAACACTATCTCTAATAAAAATACAA 210
               :::||||||   |||||||||   |||||||||:::|||   ||||||:::|||||||||
Db         124 SerGlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeuLysIleGln 143

Qy         211 AAATTAGCCAGGTGTGGTGGCACTTGCCTATAGTCCCAGCTACACAGGAGGCTGGGGCAG 270
               |||||||||      ||||||  ||||||  |||||||||      ||||||  |||
Db         144 LysLeuAlaGlyArgGlyGlyAlaCysLeu***SerGlnLeuLeuGlyArgLeuArgGln 163

Qy         271 AAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCAAGATTGCACCACTGCACT 330
               :::|||   ||||||   |||||||||||||||   |||||||||      |||||||||
Db         164 GluAsnArgLeuAsnProGlyGlyGlyGlyCysGlyGluProArgSerArgHisCysThr 183

Qy         331 CCAGCCTGGAAAAAAAGG 348
               |||||||||      |||
Db         184 ProAlaTrpAlaThrArg 189

RESULT 14
Q6ZUC0_HUMAN
ID   Q6ZUC0_HUMAN PRELIMINARY;      PRT;   166 AA.
AC   Q6ZUC0;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ43828.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Testis;
RA   Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA   Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK125816; BAC86305.1; -; mRNA.
SQ   SEQUENCE   166 AA;  17738 MW;  ED8C12B80E6D5579 CRC64;

Alignment Scores:
Pred. No.:              7.77e-24        Length:         166
Score:                  301.50          Matches:        63
Percent Similarity:     79.1%           Conservative:   5
Best Local Similarity:  73.3%           Mismatches:     17
Query Match:            32.5%           Indels:         1
DB:                     2               Gaps:           1

US-10-607-806-1_COPY_1_500 (1-500) x Q6ZUC0_HUMAN (1-166)

Qy         340 TCCAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCA 281
               ||||||||||||||:::|||||||||   ||||||||||||||   ||||||||||||
Db          61 SerArgLeuGluCysAsnGlyAlaIleSerAlaHisCysAsnLeuCysLeuProGlySer 80

Qy         280 AGTGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACC 221
               |||||||||   |||   ||||||   ||||||||||||   |||   |||||||||
Db          81 SerAspSerProAlaSerAlaSerGlnValAlaGlyThrThrGlyValCysHisHisAla 100

Qy         220 TGGCTAATT---TTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGT 164
                  ||||||  ||||||||||||:::::      ||||||||||||||||||||||
Db         101 GlnLeuIleLeuPheValPheLeuValLysThrGlyPheCysHisValGlyGlnAlaArg 120

Qy         163 CTCAAACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACA 104
               |||:::||||||   |||||||||||||||   |||   ||||||||||||||   |||
Db         121 LeuGluLeuLeuThrSerSerAspProProThrSerAlaSerGlnSerAlaGlyIleThr 140

Qy         103 GACATGAGCCACCGCACC 86
                   :::||||||||||||
Db         141 GlyValSerHisArgThr 146

RESULT 15
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
```

```
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation closely
RT   connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
CC       into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
CC       consensus sequences have been constituted that contain all six
CC       frames conceptual translations of each of these classes of Alu
CC       repeats.
CC   -!- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
CC       codon, 'XXX' is used to separate the various translation phases.
CC   -!- CAUTION: This Alu entry is provided in order to avoid the further
CC       pollution of protein sequence databases with Alu-derived amino
CC       acid sequences.
CC   -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC       primate genomes with an average spacing of 4 kb. Some of them are
CC       actively transcribed by pol III. Normal transcripts may contain
CC       Alu-derived sequences in 5' or 3' untranslated regions. However,
CC       cDNA libraries also contain partial and/or rearranged cDNAs
CC       ligated with Alu-derived sequence in any orientation. Although Alu
CC       elements (especially situated on the complementary strand) have a
CC       great potential to create additional/alternative exons,
CC       consideration should be given to the possibility that the presence
CC       of an Alu in an open reading frame may have resulted from a
CC       cloning artifact or may be due to misinterpretation of sequencing
CC       data. This point has been overlooked on several occasions, with
CC       the consequence of erroneous Alu-derived amino acid sequences
CC       being reported.
CC   -!- CAUTION: Any significant similarity of a putative protein sequence
CC       with an Alu-translated entry must be taken as a warning that a
CC       part of Alu repeat may have been artifactually included in the
CC       coding nucleotide sequence.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS; Genomic_DNA.
KW   Hypothetical protein.
FT   REGION        1     96       Frame-1.
FT   REGION      100    195       Frame-2.
FT   REGION      199    294       Frame-3.
FT   REGION      298    393       Frame-4.
FT   REGION      397    492       Frame-5.
FT   REGION      496    591       Frame-6.
SQ   SEQUENCE   591 AA;  63827 MW;  665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:                1.18e-23        Length:          591
Score:                    301.00          Matches:         56
Percent Similarity:       72.1%           Conservative:    6
Best Local Similarity:    65.1%           Mismatches:      24
Query Match:              32.4%           Indels:          0
DB:                       1               Gaps:            0

US-10-607-806-1_COPY_1_500 (1-500) x ALU1_HUMAN (1-591)

Qy        338 CAGGCTGGAGTGCAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCCAGGTTCAAG 279
              ||||||||||||||||||   :::   ||||||||||||||||   ||||||      |||
Db        506 GlnAlaGlyValGlnTrpArgAspHisGlySerLeuGlnProArgProProGlyLeuLys 525

Qy        278 TGATTCTTCTGCCCCAGCCTCCTGTGTAGCTGGGACTATAGGCAAGTGCCACCACACCTG 219
                  |||   ||||||      ||||||||||||||:::   ||||||
Db        526 ArgSerSerCysLeuSerLeuProSerSerTrpAspTyrArgArgAlaProProArgPro 545

Qy        218 GCTAATTTTTGTATTTTTATTAGAGATAGTGTTTTGCCATGTTGGCCAGGCTGGTCTCAA 159
              ||||||||||||||||||   ||||||   |||      |||   ||||||||||:::
Db        546 AlaAsnPheCysIlePheCysArgAspGlyValSerLeuCysCysProGlyTrpSerArg 565

Qy        158 ACTCCTGCTCTCAAGTGATCCACCCAACTCGGTCTCCCAAAGTGCTGGGAATACAGACAT 99
              ||||||   ||||||   |||::::::||||||||||||||||||||:::||||||
Db        566 ThrProGlyLeuLys***SerSerArgLeuGlyLeuProLysCysTrpAspTyrArgArg 585

Qy         98 GAGCCACCGCACCTGGCT 81
              |||||||||      |||
Db        586 GluProProArgProAla 591


Search completed: February  8, 2006, 09:43:15
Job time : 125.336 secs
```